ROBUST AND SECURE PERCEPTUAL IMAGE HASHING IN THE TRANSFORM DOMAIN

SUPAKORN PRUNGSINCHAI

PhD

2014

ROBUST AND SECURE PERCEPTUAL IMAGE HASHING IN THE TRANSFORM DOMAIN

SUPAKORN PRUNGSINCHAI

A thesis submitted in partial fulfilment
of the requirements of the
University of Northumbria at Newcastle
for the degree of
DOCTOR OF PHILOSOPHY

Research undertaken in the Faculty of
Engineering and Environment

NOVEMBER 2014

DECLARATION

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Name: Supakorn Prungsinchai

Signature:

Date: 22 November 2014

Acknowledgements

First of all, I am extremely thankful to Dr.Fouad Khelifi who is my PhD supervisor. He has always assisted, coached and offered me the practical guidance, useful information, recommendations and advantageous advices and possible solutions for development of my thesis. I would also like to thank Prof.Ahmed Bouridane my second PhD supervisor for his help, guidance and encouragements throughout the duration of the programme.

Secondly, I am thank all my colleagues in UK who have been supporting me and providing me with every assistance I was badly in need. I would also like to acknowledge the significant support and the perfect facilites I have received from the Department of Computer Science and Digital Technologies, Northumbria University Newcastle.

I would like to sincerely show my appreciation to Suan Dusit University (SDU); for providing me with a scholarship and financial support. Finally, I also seize this opportunity to thank my family for their great support, assistance, and encouragement during all my stay away from them.

Publications

- S. Prungsinchai, F. Khelifi, and A. Bouridane,"DCT-based Robust Image Hashing", *International Conference on Software Knowledge Information Management and Application* (SKIMA2012), China, 2012
- S. Prungsinchai, F. Khelifi, and A. Bouridane,"Sub-image based image hashing with non-negative factorisation", *IEEE International Conference on Electronics, Circuits, and System* (ICECS2012),pp.781-784, Spain, 2012
- S. Prungsinchai, F. Khelifi, and A. Bouridane,"Fourier-Mellin Transform for Robust Image Hashing", *International Conference on Emerging Security Technologies* (EST2013),pp.58-61, Cambridge, United Kingdom, 2013
- S. Prungsinchai, F. Khelifi, and A. Bouridane,"DCT sign-based Robust Image Hashing", *International Conference for Internet Technology and Secured Transactions* (ICITST2013),pp.401-405, United Kingdom, 2013

Abstract

The rapid development of multimedia devices such as computers, network technologies, and cell phones have made it easier for users to create, broadcast, convey, share, store, and distribute multimedia data including images, videos and audio files on a daily basis. However, the availability of image processing software in the public domain has facilitated illegal copying and distribution of digital images with unnoticeable quality changes. Thus, security and identification of media content has become an important and demanding area for research. Perceptual hashing is one of the recent technologies used for multimedia content security. A perceptual image hash function is a hash function that is robust against content-preserving operations (CPOs), such as noise, JPEG lossy compression and rotation.

This aim of this research is to study and investigate existing techniques and then contribute to the development of new perceptual image hashing techniques in the transform domain for image identification and copy detection applications. The design requirements for any perceptual image hashing system are robustness, discriminative capability (uniqueness), and unpredictability (security). The feature extraction stage plays a key role in ensuring the system output is robust and discriminative. This thesis mainly focuses on the robust feature extraction stage and the analysis of the proposed system's security. The following contributions have been made:

A new perceptual hashing technique using pseudo-random sub-images in the discrete wavelet transform (DWT) domain for extracting features has been developed. The idea employs a recent dimension reduction technique, referred to as non-negative matrix factorization (NMF) in the literature, for enhancing the robustness and security of the hash. This approach is proposed to select the most stable coefficients under various content-preserving operations, compact. The robust image hashes are generated by applies DWT and NMF into image. The proposed sub-images-DWT technique has been shown to yield good performance under image processing operations, but it still suffers from geometric attacks.

A new rotation-invariant FMT-based hashing technique incorporating the Fourier-

Mellin transform and using overlapping blocks to improve the robustness against rotation attacks has also been proposed. The robust FMT-based image hashing is proposed to improve its performances under rotation, translation attacks and achieve better overall robustness. The invariance property to rotation, scaling and translation of FMT makes it more suitable for image hashing. Based on our experimental results, it has been shown that the proposed FMT-based image hashing technique is robust to a large class of image processing operations and geometric attacks.

A new robust and secure DCT overlapping block-based hashing technique incorporating the discrete cosine transforms (DCT) to combat image processing attacks has been investigated. An improved DCT sign-based hashing technique robust against image processing attacks and well as small geometric manipulations developed. From the experimental results, it was observed that the low frequency coefficients for DCT sign based-image hashing were robust to a large of content-preserving operations (CPOs). The main idea was to exploit the energy compaction property of the DCT and its ability to carry information of edges and texture in DCT sign values. From the experimental results, it was observed that the low frequency coefficients for DCT sign-based image hashing were robust to a large class of content-preserving operations (CPOs). The main idea was to exploit the energy compaction property of the DCT and its ability to carry information of edges and texture in DCT sign values.

Finally, the security of the proposed image hashing systems are discussed and analysed in the light of the corresponding design requirement. The DCT sign-based image hashing scheme has hash also been shown to be the most secure technique compared to other techniques proposed in this research as it offers the highest rate of bit independence in a hash.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two Dimensional
3D	Three Dimensional
BER	Bit error rate
CCO	Content-changing operation
CPO	Content-preserving operation
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DSOC	DCT sign only correlation
DSOI	DCT sign only image
DWT	Discrete Wavelet Transform
ED	Euclidian Distance
FMT	Fourier-Mellin Transform
FPR	False Positive Rate
FT	Fourier Transform
HD	Hamming Distance
HVS	Human Visual System

NHD	Normalised Hamming Distance
NMF	Non-negative Matrix Factorisation
Non-Ov	Non-Overlapping
Ov	Overlapping
PCA	Principal Component Analysis
ROC	Receive Operating Characteristics
TPR	True Positive Rate

Chapter 1

Introduction

1.1 Background and motivation

Due to the advancement of digital devices and modern networking techniques, users can nowadays easily create, broadcast, distribute, and store digital media including images and videos daily over the social media network service such as Facebook, YouTube, and 4shared etc. Because of the easy-to-copy nature of digital media, digital data can be illegally distributed or forged this threatening data security and integrity. Therefore, measures for security should be considered as follows:

- ***Content authentication***: This is because digital multimedia can easily be manipulated or the content can be tampered with. For example, objects can be removed or added in an image easily by using image processing tools. Therefore, the aim of content authentication is to verify the integrity of digital image data and identify malicious attacks has become one of the most important issues in digital media security.

- ***Copyright protection***: Users upload their images/video into public websites, everyone could download without any authorisation. The goal of copyright protection is to identify perceptually identical even if they undergo different types of distortion induced by the imperfect transmission channel (see Figure 1.2) or protection

of intellectual property right from illegal/unauthorised usage of these digital media data.

Thus, effectively protection of the copyright of digital media data has become a real issue which needs to be resolved. There are two approaches for the intellectual property protection of digital media; watermarking (Cox et al., 1996; Yeung and Mintzer, 1997; Wu and Liu, 1998; Lu and Liao, 2001; Xie and Arce, 2001; Lu et al., 2003) and digital signature (Schneider and Chang, 1996; Lin and Chang, 1998; Venkatesan et al., 2000; Xie and Arce, 2001; Hampapur and Bolle, 2001; Lin and Chang, 2001; Mihçak and Venkatesan, 2002; Monga and Evans, 2004; Mucedero et al., 2004; Swaminathan et al., 2006; Sunil and Yoo, 2008; Khelifi and Jiang, 2010; Zauner, 2012).

Watermarking has been mainly developed in order to authenticate the integrity of media data and protect digital copyrights. The fundamental idea of watermarking is to embed invisible secondary data called watermarks depending on the application, onto the digital images or videos prior to distribution. Therefore, all copies of the marked content contain the watermark, which can be extracted as identification information to prove ownership. Watermarking has been widely used for the following purposes: broadcast monitoring, copyright protection, copy control, authentication and proof of ownership (Cox et al., 2000). One of the limitations of watermarking is that the embedding process would inevitably cause slight changes upon the media content, especially when the embedded watermark signals are required to be robust against signal processing attacks. Therefore, a balance between the strength of embedded watermark signals and the content quality of the host media data needs to be considered.

The digital signature is a set of features extracted from the media itself, which can represent the content of the original data. This means the digital signature does not require additional information, just a media itself. Generally, media data such as images and videos contain enough unique information to be used as a content signature or as content identification (content ID) for detecting copies, especially those that have been illegally distributed. For example, the owner of the film *Fast and furious* created a set of video signatures (feature vectors), which meant that they could be used to rapidly find the movie

clips. If the owner suspects that it is being illegally distributed on the internet, they can use a copy detection system. The copy detector investigating video copies in a database, in which videos are collected from the web pages.

Classic cryptographic hash functions, e.g., MD5 and SHA-1, or message authentication codes, convert an input message (document, image, video, text, etc.) into a fixed-length bit string. Cryptographic hash functions are typically used for as digital signature to authenticate the message being sent, therefore, the recipient can verify its source. In general, authentication means deciding whether an object is authentic or not. That is, if every single bit in the transmitted data is matched in the original object. Thus, cryptographic hash functions are suitable for such tasks. However, they are extremely sensitive to single-bit changes in the input data and are not suitable for digital multimedia, especially digital images; where perception of the data is needed. This is because in real applications digital images often undergo signal processing operations, such as JPEG lossy compression, noise, rotation and image enhancement. These signal processing operations would definitely change the binary representation of the input image and hence perceptual hashing would be the solution. With this in mind, the perceptual image hashing concept is presented and as an efficient tool to address issues of image copyright protection.

1.2 Perceptual image hashing

1.2.1 Concept and properties

As an alternative way to ensure efficient image copyright protection, perceptual image hashing has been proposed to generate a robust, unique, and secure feature for each image and thus, calculate the hash values of these features, without any watermark embedded in host images. Authentication/identification of an image is performed through comparison of hash values of original data and the query data using specific functions. The image hash depends on the image content itself.

Let I denote a particular image and $\hat{I}$ be a modified version which is "***perceptually***

similar" to I and J denotes an image that is "***perceptually different***" from I. Let θ_1,θ_2 two positive values that satisfy $0 < \theta_1, \theta_2 < 1$. An hash function denoted by $H(.)$, produces a hash signature of fixed length depending on a secret key, K. The properties of a perceptual image hashing function are identified as follows:

1. Robustness:

$$(H(I, K) \approx H(\hat{I}, K) \geq 1 - \theta_1, 0 \leq \theta_1 \leq 1 \tag{1.1}$$

The robustness property requires for any pair of perceptually similar/identical images should have similar hashes. While for digital image data, perceptually insignificant distortions introduced to original images due to compression or noise transmission channels or via Internat. Therefore, it is main required to guarantee that perceptually similar images hash similar image hashes, and image hashes should be robust enough to such content-processing operations (CPOs) attacks for identification purpose. An example is illustrated in Figures 1.2(a) to (g), which includes the original image and its distorted copies under different distortions such as JPEG lossy compression, additive white Gaussian noise (AWGN), median filter, histogram equalisation, rotation and translation. Perceptually, these images are identical in human visual system (HVS). The perceptual robustness of image hashing guarantees that these images have very similar hashes, if the system is robust enough against these attacks.

2. Discriminability:

$$(H(I, K) \neq H(J, K) \geq 1 - \theta_2, 0 \leq \theta_2 \leq 1 \tag{1.2}$$

The discriminability of the image hashing system guarantees that perceptually distinct images should have different hashes.

3. Unpredictability:

$$(H(I,K)); f_h(1) \approx f_h(0) \approx 0.5 \tag{1.3}$$

Where $f_h\ (x)$ is the probability mass function for h. With this property the hash values should be approximately equally distributed. Security is an important concern for image hashing. Pseudo-randomisation techniques are incorporated into the image hash generation process to enhance the security of image hashes by using secret keys.

4. Compactness:

$$Size(H(I,K)) \ll Size(I); \tag{1.4}$$

The size of the image hashes should be much smaller than the original image I. The compact image hashes should facilitate the search process in a database of hashes and should require less storage space (Yang and Rhee, 2010).

1.2.2 Perceptual image hashing framework

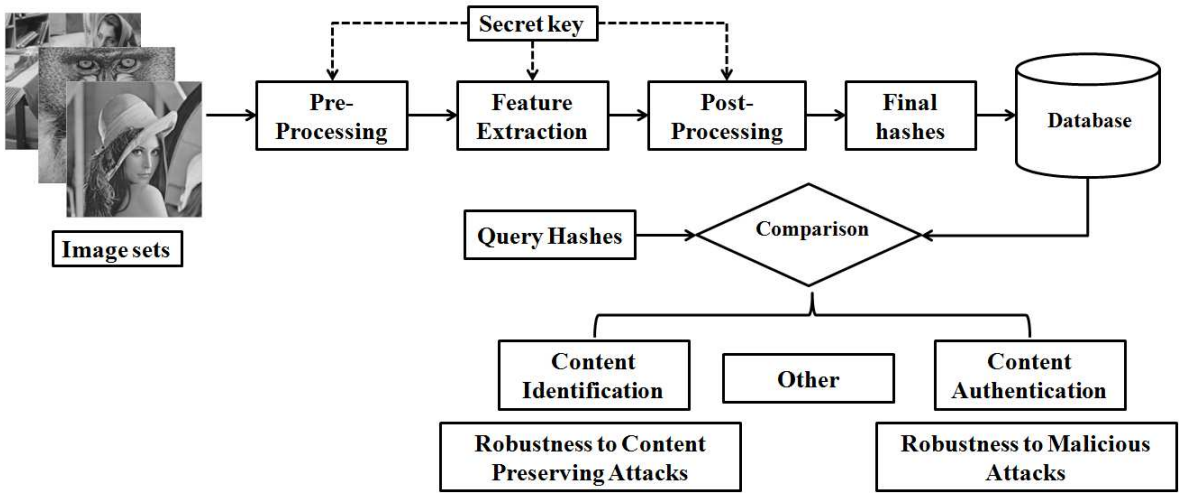


Figure 1.1: The generic framework of image hashing: Adapted from Yang and Rhee (2010)

A perceptual image hashing system, as illustrated in Figure 1.1, generally consists of three main stages:***pre-processing***, ***feature extraction*** and ***post-processing***. The robustness of image hashing arises from robust feature extraction and the post-processing mainly contributes to the final hash. The security of image hash can be cooperated into one or two

of the three steps by using a secret key. After reviewing digital image hashing algorithm design, the design framework of digital image hashing depends on different application scenarios. For example, if the digital image hashing algorithm is designed for content identification or copy detection purposes, it mainly concentrates on the robustness against content-preserving operations (CPOs) that do not destroy the perceptual quality of the image. However, for image authentication purpose, it mainly focuses on the malicious attacks (content-changing operations (CCOs)) in the image content such as removal and object insertion, image hashes should be sensitive to these perceptually significant attacks.

(a) Original “Lena” image

(b) JPEG lossy compression

(c) AWGN

(d) Median filter

(e) Histogram Equalisation

(f) Rotation

(g) Translation

Figure 1.2: Examples of distorted “Lena” image copies under different content-preserving operations (CPOs)

Pre-processing

The pre-processing step aims to decrease the sensitivity of feature extraction against minor distortions on the image such as noise and lossy compression, before the feature extraction step. The common pre-processing operations applied on the images are listed as follows:

- ***Filtering*** (Swaminathan et al., 2006; Roy and Sun, 2007; Xiang et al., 2007): This is an efficient method to improve the robustness of the extracted features against noise. The popular filters, such as Gaussian filter, could be applied on image hashing to assist with noise reduction and also eliminate some details of the image contents and generate blurred images. Thus, the image hashing scheme is robust enough for blurring distortions.

- ***Illumination normalisation*** (Fridrich and Goljan, 2000; Swaminathan et al., 2006): This is the way to improve the extracted feature against a change in brightness or gamma correction attacks. Illumination normalisation processes such as histogram equalisation can effectively render the extracted features invariant to illumination changes.

- ***Resizing*** (Mihçak and Venkatesan, 2002; Monga and Evans, 2006; Swaminathan et al., 2006; Lv and Wang, 2009): This is the way to improve the efficiency of hash generation. The features extracted from the image with a standardised size are more robust against the aspect ratio change.

- ***Colour space dimension reduction*** (Fridrich, 2000; Venkatesan et al., 2000; Mihçak and Venkatesan, 2002; Kim, 2003; Lefebvre et al., 2003; Lu and Liao, 2003; Monga and Evans, 2006; Swaminathan et al., 2006; Lin et al., 2007; Brasnett and Bober, 2008; Lv and Wang, 2009): This is a common operation applied in most digital image hashing algorithms to reduce the computational cost for feature extraction (e.g. 3D to 2D).

Feature extraction

Robust feature extraction is the primary goal of perceptual image hashing algorithms. The extracted features are based on the characteristics of images which should be unique and distinctive enough for content identification. Hash values are expected to survive or robust on insignificant content-preserving operations (CPOs) as long as two images appear perceptually similar to the human visual system (HVS). Thus the hashes should be as similar as possible. Likewise, two images that are perceptually different should correspond to dissimilar hashes. Hence, robust feature extraction is a key step in image hashing due to the critical robustness requirement. Based on the literature review and my own study, the feature extraction operations that are applied on the image to search for certain features to resist some types of image distortions are classified as follows:

- ***Techniques based on statistical information*** (Motwani and Raghavan, 1996; Schneider and Chang, 1996; Venkatesan et al., 2000; Kailasanathan and Nani, 2001; Kim, 2003; Wu et al., 2007; Xiang et al., 2007; Zou et al., 2009): This group of techniques extracts hash features by calculating the images statistics of pixels values of the image in the spatial domain, such as image intensity, mean, variance and other higher order moments. The statistic features are more robust than the raw pixel values against noise and compression distortions, although with less distinctiveness.

- ***Techniques based on low-level features*** (Fridrich, 2000; Lu and Hsu, 2005; Monga and Evans, 2006; Roy and Sun, 2007; Zou et al., 2009) : This group of processes extracts local feature patterns usually by including edges, interest points, corners, blobs, etc. The advantage of applying local features is mainly for robustness against geometric attacks, however, it is sensitive to noise addition, blurring, compression attacks.

- ***Techniques based on dimensionality reduction*** (Fodor, 2002; Kozat et al., 2004; Monga and Mihçak, 2007; Tang et al., 2008; Lv and Wang, 2008; Hassan et al., 2012; Tang et al., 2013): This group of approaches extracts hash features by a process of reducing a high dimensional parapets into a relatively low dimensional

data as well as maintaining the properties of the original data. The advantage of applying dimensionality reduction techniques are mainly for robustness against noise addition, blurring and compression attacks. However, their performance under large geometric attacks is still limited.

- ***Techniques based on invariant properties in transformed domains*** (Fridrich and Goljan, 2000; Venkatesan et al., 2000; Lin and Chang, 2001; Lefbvre et al., 2002; Mihçak and Venkatesan, 2002; Kailasanathan et al., 2003; Kim, 2003; Lu and Liao, 2003; Kozat et al., 2004; Seo et al., 2004; Lu and Hsu, 2005; Monga et al., 2005; Harmanci et al., 2006; Swaminathan et al., 2006; Guo and Hatzinakos, 2007; Roy and Sun, 2007; Gerold and Andreas, 2008; Lv and Wang, 2009): This group of techniques extracts the hash features by transforming the image from the spatial domain into the transform domain. The coefficients in the transform domain can be critical features and robust enough against a large class of image processing operations and attacks. The transforms used to extract features include discrete wavelet transform (DWT), discrete cosine transform (DCT), fourier-Mellin transform (FMT), Radon transform (RT), etc.

Post-processing

This step is concerned with compactness in image hashing, which is another critical property of image hashing algorithms. Robust features need to be compressed into a short real-valued or binary sequence of fixed-length, which can be considered to be a "dimension reduction process". Some typical techniques are summarised as follows:

- ***Quantisation*** (Fridrich, 2000; Venkatesan et al., 2000; Kailasanathan and Nani, 2001; Lin and Chang, 2001; Mihçak and Venkatesan, 2002; Kailasanathan et al., 2003; Kim, 2003; Lu and Liao, 2003; Seo et al., 2004; Lu and Hsu, 2005; Monga, 2005; Swaminathan et al., 2006; Guo and Hatzinakos, 2007; Lin et al., 2007; Roy and Sun, 2007; Xiang et al., 2007; Brasnett and Bober, 2008): Here, continuous feature space is converted to finite discrete feature space. The popular approaches include

interval quantisation, binary quantisation using ordinal measures or threshold for image hashing generation.

- ***Compression and coding*** (Venkatesan et al., 2000; Norcen and Uhl, 2005; Swaminathan et al., 2006; Lin et al., 2007): These techniques are used in communications and can be applied to compress the robust features into short image hashes. Popular techniques include distributed source coding (e.g., Wyner-Ziv, Slepian-Wlof), and Error-Correcting Coding (ECC) etc.

- ***Random projection*** (Harmanci et al., 2006; Swaminathan et al., 2006; Lin et al., 2007; Monga and Mihçak, 2007; Roy and Sun, 2007; Lv and Wang, 2008, 2009): The random projection approach can offer a performances comparable to that of the conventional dimension reduction methods, for instance the Principal Component Analysis (PCA). Random projection is a pseudo-randomisation process that can enhance the security of the designed image hashing scheme.

- ***Clustering*** (Kailasanathan and Nani, 2001; Kim, 2003; Monga and Evans, 2006; Roy and Sun, 2007): This method divides the feature space and map similar features into the same centroid of clusters.

Security incorporation

A further important property of image hashing is "***security***". The basic idea is to make image hashes unpredictable by incorporating a secret key into the hash generation to make it as a pseudo-randomisation process. A change in the secret key should significantly change the hashes to ensure that users cannot guess or generate the right hash of an image without the correct secret key. Therefore, with the secret key, the security of image hashes can be controlled by approved users and prevent unauthorised access, which facilitates the application of copyright protection. The way to security incorporation can be viewed as follows:

- ***Randomised tiling*** (Venkatesan et al., 2000; Mihçak and Venkatesan, 2002; Kozat et al., 2004; Meixner and Uhl, 2005, 2006; Monga and Evans, 2006; Monga and Mihçak, 2007; Tang et al., 2008; Lv and Wang, 2009; Jie, 2013): Images are randomly partitioned into overlapping sub-images based on the selected secret key. These sub-images can be circles, or rectangles, squares with a selected radii or size. Subsequently, the extracted features from these randomised sub-images enhance the security of the final image hash. Typically, randomised tiling is applied in the pre-processing step. Nevertheless, it should be noted that random partitions are sensitive to geometric attacks.

- ***Randomised projection*** (Schneider and Chang, 1996; Dittmann et al., 1999; Venkatesan et al., 2000; Mihçak and Venkatesan, 2002; Harmanci et al., 2006; Swaminathan et al., 2006; Kitanovski et al., 2007; Lin et al., 2007; Monga and Mihçak, 2007; Roy et al., 2007; Xiang et al., 2007; Lv and Wang, 2009): This method is applied in both the post-processing and compression steps to project robust features into a lower dimension based on the projection matrix, whose entries are random variables determined by the selected secret key.

- ***Randomised transform*** (Fridrich and Goljan, 2000; Meixner and Uhl, 2006; Lin et al., 2007; Tang et al., 2008; Lv and Wang, 2009; Fawad et al., 2010): After extracting the robust features, another randomised domain determined by the selected secret key is further used to make the features unpredictable. It is inherently a pseudo-encryption process applied in the feature extraction step.

- ***Traditional cryptography*** (Lin and Chang, 2001; Xie et al., 2001; Lu and Liao, 2003): Cryptography could be employed for encrypting features after the compression step, although they are sensitive to a single change of encrypted data.

1.2.3 Comparison and decision making

The distance metrics and classifiers reviewed in this subsection are mainly applied to evaluate the robustness and discriminative capability of image hashing schemes. Ideally, a

robust and secure image hash is generated and saved in the database, as an index. When a query hash is received, it will be compared by calculating the distance between the hashes, and as a result, the outcome is a "***similarity score***". The distance metrics that are selected to measure the similarity between hashes to make decisions, are very important in image hashing.

Distance metrics

The choice of distance metrics depends on the type of hashes. Given two image hashes $H_1 = h_1(1), h_1(2), \ldots, h_1(n)$ and $H_2 = h_2(1), h_2(2), \ldots, h_2(n)$ of two images I_1 and I_2 with hash length n, the following distance metrics are usually employed:

Hamming distance (HD):

The Hamming distance is utilised to measure the similarity between two binary hash vectors by comparing with the bit-by-bit. The Hamming distance is given as:

$$HD(H_1, H_2) = \sum_{i=1}^{n} |h_1(i) \oplus h_2(i)| \tag{1.5}$$

Normalised Hamming distance (NHD):

The Hamming distance can be normalised with respect to the length n of the binary strings, then normalised in the [0,1] range. The two images are perceptually similar and the distance is close to 0, whereas the distance is expected to be close to 0.5 for two distinct images. The normalised Hamming distance is defined as:

$$NHD(H_1, H_2) = \frac{1}{n} \sum_{i=1}^{n} |h_1(i) \oplus h_2(i)| \tag{1.6}$$

Euclidean distance (ED):

The Euclidean distance is a technique that is suitable for non-binary vectors (real-values or integers). It is defined as the square root of the sum of the squares of the differences

between the corresponding hash values.

$$ED(H_1, H_2) = \sqrt{\sum_{i=1}^{n}(h_1(i) - h_2(i))^2} \tag{1.7}$$

Bit Error Rate (BER):

The bit errors rate ρ as the number i of bit errors of the hash normalised by the length n of the hash:

$$\rho := \frac{i}{n} \tag{1.8}$$

where $i \in \{0, 1, \ldots, n\}$ and $0 \leq \rho \leq 1$.

The number of the bit errors i equals the hamming distance of the hash values. Therefore, a perceptually similar image should yield a *BER* close to 0.

1.2.4 Classifiers

Based on the similarity between two hashes, computed by the selected specific distance metrics, a similarity score is output. A classifier is employed to make a decision in connection with the content identification. The classifier is defined as:

$$Distance((H_1), (H_2)) \leq \tau, \tag{1.9}$$

when I_1 is similar to I_2

$$Distance((H_1), (H_2)) > \tau, \tag{1.10}$$

when I_1 is different from I_2, where τ is the selected threshold.

1.2.5 Receiver Operation Characteristics (ROC) curves

ROC graphs are two-dimensional graphs defined by the True Positive Rate (TPR) which is plotted on the y-axis and False Positive Rate (FPR), which is plotted on the x-axis. The ROC graphs depicts relative trade-off between the true positive (benefits) rate and false

positive (cost) rate of the system (Egan, 1975; Swets et al., 2000; Fawcett, 2006). Figure 1.3 shows an ROC graph with five classification labeled A through E. The diagonal divides the ROC space, with the points above the diagonal represent high-quality classification results, and the points below the line signifying poor classification results. Therefore, the best possible prediction method would produce a point in the upper left corner or coordinates (0,1) of the ROC space (Zweig and Campbell, 1993).

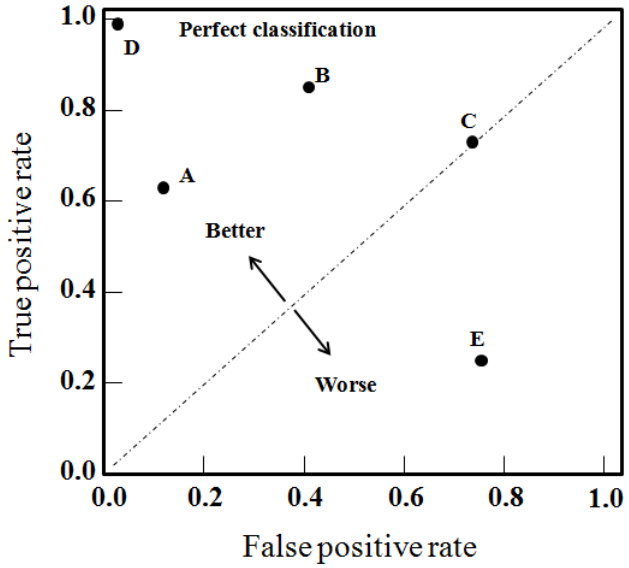


Figure 1.3: Example ROC curve analysis

1.3 Aim and Objectives

This aim of this research is to investigate the performance of existing techniques and contribute to the development of new techniques of perceptual image hashing to be used for image identification. The main objectives are as follows:

1. To investigate the performance of Discrete Cosine Transform (DCT), Discrete Wavelet transform (DWT), and Fourier-Mellin Transform (FMT) for improving feature extraction.

2. To develop new techniques of perceptual image hashing for content identification

based on the methods and framework of objectives 1.

3. To evaluate the performance of the perceptual image hashing algorithms including a comparative study alongside existing and related techniques: this will be carried out by considering content-preserving operations (CPOs) through intensive experiments.

1.4 Thesis contributions

Based on the four basic requirement properties of image hashing: robustness, discriminability, compactness, and security. It is obvious that the key point in designing desirable digital image hashing algorithms are feature extraction, feature compression and the security incorporation. In this thesis, I will focus mainly on the robust feature extraction stage and security issues. The contributions of this thesis are as follows:

1. To complete a literature review on image hashing including an analysis and comparison of various robust feature extraction methods, and research into the future direction of image hashing.

2. To introduce a robust and secure perceptual image hashing technique based on pseudo-random sub-images in the DWT domain for extracting features. The idea also makes use of a recently proposed dimension reduction technique, referred to in the literature, as "*Non-negative Matrix Factorisation*"(NMF) for enhancing the robustness of the hash to achieve superior identification performances under various distortions and attacks. Experimental results show that the proposed image hashing scheme can provide an improved performance under image signal operations, although it can still suffer from geometric attacks.

3. Propose using the popular rotation, translation and scaling invariant feature transform by using the FMT domain and overlapping blocks to improve the security and robustness against geometric attacks. This image hashing concept develops the performance of the image hashing technique against rotation and translation attacks.

Experimental results show that the proposed image hashing scheme is robust in wide range of distortions.

4. To propose a robust and secure DCT-based image hashing method and DCT sign-based image hashing method. The DCT is well-known for its compression capability; therefore, it is widely used in image compression standards. The DCT coefficients were utilised as features, to identify images under various distortion attacks. Experimental results demonstrated that the DCT sign-based image hashing scheme offers an excellent robustness against signal processing operations and geometric attacks, especially translation and outperforms the conventional DCT overlapping block-based image hashing scheme.

1.5 Review of related work

As mentioned earlier most of the existing perceptual image hashing algorithms are generated from significant features that represent the image semantic content and use them during identification or authentication. This can be classified into four types as follows:

1.5.1 Techniques based on statistic information

The statistical information, such as image intensity, mean and variance, is invariant under small perturbations to the image. The early research of Schneider and Chang (1996) used an intensity histogram of image blocks to create image hash values. The histograms are encrypted by using pubic key to obtain the final image hash, this then needs to be stored and decrypted again for verification. In the verification step, the Euclidean distance between intensity histograms are used as a measure to verify the image. The most significant drawback of the scheme is that it is easy to attack the image without altering its histogram. Venkatesan et al. (2000) proposed an image hashing method that was used for indexing and database searching. This scheme is based on an image statistic computed from randomised rectangles in the variant sub-bands, in a wavelet decomposition of the

image. In the algorithm, a wavelet decomposition of the image is computed first and each of the sub-bands is randomly divided into rectangles by using a secret key. Each rectangle's statistics are calculated and quantization is used by applying a randomised rounding (Motwani and Raghavan, 1996) ensures that the final hash values, as binary strings are random. The quantised statistics are the decoding stage of the Reed-Muller Error correcting code (Blahut, 1983) to generate the final hash value. This technique has been shown to be robust against image processing operations and geometric attacks. Kailasanathan and Nani (2001) proposed an image hashing scheme based on K-means segmentation to extract statistics such as variance, mean and other higher order moments from image blocks to use in image hashing.

Furthermore, Xiang et al. (2007, 2012) proposed a robust image hash algorithm by using the histogram shape invariance that is robust against geometric attacks and image processing operations. This scheme does achieve a satisfactory robustness performance for most signal processing operations and geometric attacks. Tang et al. (2012) proposed a robust image hashing algorithm using an advantaged histogram of colour vector angles to generate hashes. This scheme is robust against rotation with an arbitrary degree. The advantage of the statistic information based method is its robustness against perturbations to the image. The schemes in Schneider and Chang (1996); Venkatesan et al. (2000); Xiang et al. (2007) achieve robust capability under geometric operations. Nevertheless, the security of this kind of method is very weak, but random partitioning may solve the security problem under the assumption that the sizes and the positions of image blocks are secure enough against attackers.

1.5.2 Techniques based on dimensionality reduction

Dimension reduction is the process of reducing high dimensional datasets into relative low dimensional datasets, as well as maintaining the properties of the original data (Fodor, 2002). Kozat et al. (2004) suggested viewing images and attacks on a sequence of linear operators and proposed novel hashing algorithms based on Singular Value Decomposition (SVD). This scheme, called SVD-SVD hashing, first applies SVD on the image to extract

intermediate features. Following this, a secondary image is constructed from the intermediate features and decomposed by SVD again to obtain the final hash. The SVD-based image hashing algorithm is robust against severe geometric attacks on images; however its discriminative capability between different images needs to be improved.

Monga and Mihçak (2007) proposed the use of non-negative matrix factorisation (NMF) to derive image hashing because of its non-negativity constraints. This proposed called NMF-NMF hashing, is applied NMF to sub-images. The combination coefficients in matrix factorisation are used to construct a secondary image, and subsequently its low-rank matrix approximation is obtained by using NMF again, in order to generate the secure hash sequence. The two-stage cascade application of NMF on images obtains a higher robustness under content-preserving operations (CPOs) while reducing the misclassification rate for the images under content-changing operations (CCOs). Lv and Wang (2008) presented image hashing based on a Fast Johnson-Lindenstrauss Transform (FJLT). This approach has a comparable and robust capability as a NMF-based scheme, although it has a lower computational cost. Tang et al. (2008) observed the invariant relation that exists in the NMF coefficient matrix and used this property to construct robust hashes. This scheme is robust against common signal processing operations such as JPEG lossy compression, additive noise and watermarking embedding. However it is, fragile to image rotation. Hernandez et al. (2011) proposed an image hashing scheme by using normalisation and SVD decomposition hash functions to generate a hash value. This scheme applied an image normalisation technique on randomly selected sub-images, as a pre-processing step aimed at increasing robustness against rotation, scaling and JPEG compression. The first SVD decomposition function is applied to each sub-image. Subsequently, rearranging the matrices and again applying an SVD decomposition function generates the final binary hash value. Their results were a significant improvement in terms of the Hamming distance against some image processing operations and geometric attacks such as JPEG lossy compression, rotation and scaling. Hassan et al. (2012) presented secure and robust image (visual) hashing based on the DWT and NMF. This scheme first applied a low pass filter and histogram equalisation, as a pre-processing step, then applied DWT decomposition

on the image. NMF is applied on low-frequency coefficients to defer the coefficients and then to generate the final binary hash value. From their results, they proposed a method that can achieve robustness against perceptually insignificant manipulations and has an enhanced performance in terms of local tampering detection. Tang et al. (2013) proposed robust perceptual image hashing based on ring partition and NMF. The key is a novel construction of the rotation-invariant secondary image which helps to make the image hash resistant to rotation and has a desirable discriminative capability. This scheme obtains a satisfactory robustness against the geometric operation, and discriminative capability is ensured. The image hashing generated by dimension reduction methods depends on the creation of an image on an adaptive basis, which is an unsupervised learning process. Therefore, a trade-off between the efficiency and classification performance would need to be considered when designing an image hashing algorithm via a dimension reduction technique.

1.5.3 Techniques based on low-level features

The low level features are edges or interest points information in the image. Robust low level feature extraction is the main task for this kind of approach. The properties of the image hashing algorithm based on low level features mainly depend on the performance of low level feature detections. In their earlier works, Monga and Evans (2004, 2006) exploited the end-stopped wavelet cells to detect visually significant feature points. Based on this evaluation and comparison results of other feature points detectors, the end-stopped wavelet gained robustness with the content-preserving operations (CPOs). In this scheme the feature points are extracted by using end-stopped wavelet transform. To make a short hash, an iterative algorithm proposed in Mihçak and Venkatesan (2002) is employed to obtain a hash value. The image hashing based on feature-points detection may be fail in the case of a smooth texture on image. Lu et al. (2004) and Lu and Hsu (2005) proposed robust hashing for copy detection and tracing images. The Harris detector is applied on the image to detect robust points. Then, Delaunay tessellation is performed using the obtained points to generate the set of meshes. There is no secret key in the scheme, thus,

the security cannot be evaluated and the mesh normalisation during hash generation is complex. In addition, it has a significant time cost. Roy and Sun (2007) presented an image hashing scheme for detecting and localising image tampering. The localisation of the tampering is an optional functionality of image hashing. This scheme consists of two parts: the first part is used for authentication only, and is this part based on the Scale-Invariant Feature Transform (SIFT) proposed in Lowe (2004), which is robust to several geometric transforms. The second part is used for tampering localisation and is based on local histograms of directions of the edges. The performance of both the robustness and collision resistance are dependent on parameters which are analysed in Roy et al. (2008). The properties of low-level features are mainly dependent on the performance of a low level feature detector. It must be robust enough against content-preserving operations (CPOs) and have sufficient discriminative capability for content-changing operations (CCOs), and moreover, an image hashing algorithm must be constructed under a suitable key projection. Additionally, the parameters used to construct the hash, such as the number of feature points, the edges and the number of meshes should be selected carefully in order to avoid unnecessary costs.

1.5.4 Techniques based on invariant properties in transformed domains

The invariant in the transformed domain is the image from the spatial domain into other domains such as DCT, DWT, or DFT. Different algorithms utilize different invariant properties in various domains to construct the robust image hashes. Early research proposed by Fridrich and Goljan (2000) constructed the image hash by selecting DCT coefficients. This method is based on a large absolute value of low-frequency coefficients on the zero-mean random smooth patterns based on a secrete key. The hash extracted via this method is fairly robust to slight content-preserving operations (CPOs). The weakness in this method is that it is very sensitive to a small angle image rotation. Mihçak and Venkatesan (2001) proposed an image hashing algorithm based on the DWT coefficients by using

iterative approach to binarise the DC-subband (LL subband) of a 3-level Haar wavelet decomposition of the image. During the iterations, the significant features are preserved, whereas insignificant features are eliminated. Lin and Chang (2001) presented a robust image authentication system based on invariant relations between DCT coefficients at the same position in separate blocks. This image hashing method can prevent malicious manipulations, but allows JPEG compression and is fragile to rotation. Lefbvre et al. (2002) were the first to use Radon transform (RT) to construct robust hashes. The Radon transform, which is largely used in magnetic resonance imaging, is also robust against image processing basic attacks. Seo et al. (2004) presented a new method for image fingerprinting using auto correlation of each projection in the Radon transform domain to make image hashing robust against affine transformations. Swaminathan et al. (2004) introduced FMT to image hashing, the proposed image hash is resilient to geometric and filtering operations, and is secure against guessing and forgery attacks. Fawad and Siyal (2005) proposed a secure and robust hashing scheme for image authentication by using the property of DWT and SHA-1. These methods allow acceptable manipulations like JPEG compression and low pass filtering and are sensitive enough to detect malicious manipulations. The security is achieved by using the permutation key in the feature extraction stage and by encrypting the final hash value using the sender's private key, an attacker can not easily calculate the image hash using the DWT without the private key. Roover et al. (2005) introduced a robust video hashing algorithm based Radial Hashing (RASH) algorithm. They did this by dividing an image into a set of radial projections of image pixels, then extracting a Radial Variance (RAV) vector from these radial projections and a compressed vector by DCT. The RASH algorithm is resilient to geometric attacks, (like image rotation and re-scaling), but its discriminative capability needs to be improved. Swaminathan et al. (2006) used the DFT coefficients to produce image hashing. This scheme is resilient to several content-preserving operations (CPOs), such as compression, filtering, and common geometric operations up to 10° of rotation and 20% of cropping. This scheme also has good discriminative capabilities and can identify malicious manipulations such as a cut-and-paste type of editing. Hadmi et al. (2010) analysed the robustness of wavelet-based perceptual signatures. This was

achieved by generating signatures from the coefficients of the subband LL after a DWT transformation. The proposed method is efficient, robust against common content preserving manipulations. Lei et al. (2011) performed the Radon transform (RT) on the image and calculated the moment features, which are invariant to translation and scaling in the projection space. The significant DFT coefficients of the moments are used to produce the image hash bits. This RT-DFT scheme for image hashing produced superior results than Seo et al. (2004) and Swaminathan et al. (2006) techniques in terms of perceptual robustness and discriminative capability. Recently, Jie (2013) proposed a block-DCT and Principal Component Analysis (PCA) based image hashing algorithm. The main idea of this algorithm was to integrate a colour histogram and DCT coefficients of image blocks, then to compress robust features as inter-feature with PCA, and subsequently muse a threshold to create a robust hash. The most widespread frequency transforms to dissimulate data in images are the DFT, DCT, and DWT. However, most of the transforms are not new and the coefficients here may be easily attacked (Fawad and Siyal, 2006). Thus, improving the security may sacrifice somewhat of the robustness to the content-preserving operations (CPOs).

1.6 Thesis outline

The thesis outline is summarised as follows:
Chapter 1 provides the introduction of this thesis, which includes the overview, challenges and motivations for this work. The review of related work is then presented. Next, the principles of perceptual image hashing, which is used in this work, are introduced. In addition, the aims and objectives, thesis distribution and thesis outlines are also presented.
Chapter 2 presents an introduction to image representation in the transform domain. Then, the several transforms used in this research work, such as DCT, FMT, and DWT, are illustrated. The chapter end with a conclusion.
Chapter 3 presents the proposed framework for perceptual image hashing in the DWT domain. This includes illustrating the approach, random partition, NMF and the identifi-

cation and evaluation experiments. The experimental results are then illustrated. Subsequently, the conclusion is drawn.

Chapter 4 presents the proposed framework for perceptual image hashing in the FMT domain, which includes a description of the technique with experimental assessments and analysis. A conclusion is provided on the performance of the system.

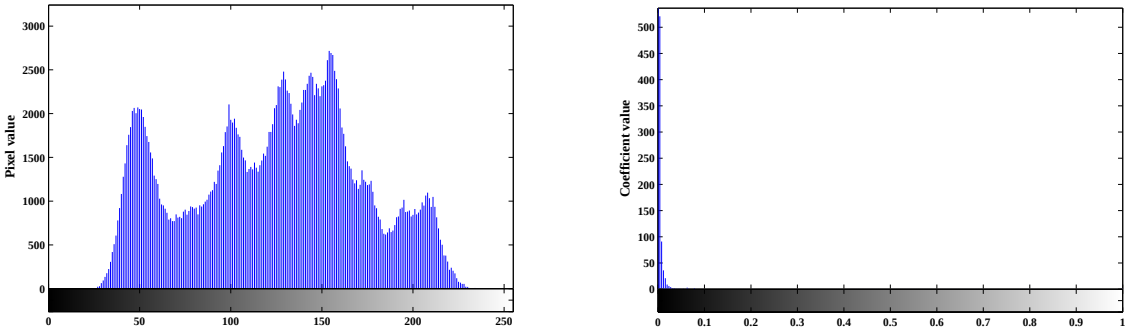
Chapter 5 presents the proposed framework for perceptual image hashing in the DCT domain, which contains of two techniques: DCT overlapping block-based image hashing and DCT sign-based image hashing. Experimental results are then illustrated and discussed.

Finally, chapter 6 draws to a conclusion to the entire thesis by making a summary of the main contributions and presents recommendations for futures work.

Chapter 2

Image representation in the transform domain

An image transform can be thought of as a way to change the statistics of the source image. The most attractive properties exhibited in the transform domain are energy *compaction* and *decorrelation*. Energy compaction suggests that most of the signal information tends to be concentrated in a few low frequency components of the transform. Figure 2.1 shows the energy distribution of a standard image "Lena" in both spatial and DCT domains. Decorrelating transforms remove linear dependencies from the data. Therefore, a set of components are produced such that, when they are individually quantised and entropy coded, the resulting symbol stream is reduced substantially, when compared to applying the quantisation directly on the image data. In fact, in the spatial domain, a portion of information carried in a given pixel may also exist within another adjacent pixel. As mentioned earlier, this is referred to as spatial redundancy. Redundancy reduction aims at removing duplication from the image data. In the following, the most widely used decorrelating transform are reviewed.



(a) (b)

Figure 2.1: Histogram of the image "Lena" image, (a) Spatial domain (b) DCT domain Only a portion ($\frac{1}{64}$) of the entire transform image is taken at the top-left side, where the magnitude of the spectrum is normalised in [0,1] and displayed to show the behaviour in the lower frequencies

2.1 Discrete Fourier Transform (DFT)

The Fourier Transform (FT) is an important image processing tool which is used to decompose a signal into its sine and cosine components (Bracewell, 1999). It is generally a complex valued function which is defined for an integrable function x as

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-i2\pi ft}dt, f \in \Re. \tag{2.1}$$

When the independent variable t represents time, the transform variable f represents ordinary frequency. In the Fourier domain signal, each point represents a particular frequency contained in the time domain signal. The signal $x(t)$ can be reconstructed from $X(f)$ by the inverse transform

$$x(t) = \int_{-\infty}^{+\infty} X(f)e^{-i2\pi ft}df, t \in \Re. \tag{2.2}$$

The interpretation of $X(f)$ is aided by expressing it in polar coordinate form as

$$X(f) = |X(f)|e^{i\Phi(f)} \tag{2.3}$$

where $|X(f)|$ and $\Phi(f)$ represent the amplitude and the phase of $X(f)$, respectively. Let $x(t) \Leftrightarrow X(f)$ denotes that $x(t)$ and $X(f)$ are a Fourier transform pair. Some important properties of the Fourier transform are

- Linearity

$$ax_1(t) + bx_2(t) \Leftrightarrow aX_1(f) + bX_2(f)$$

- Convolution

$$x_1(t) * x_2(t) \Leftrightarrow X_1(f)X_2(f)$$

- Scaling

$$x(at) \Leftrightarrow \tfrac{1}{|a|}\ X\ (\tfrac{f}{a})$$

- Time shift

$$x(t - t_0) \Leftrightarrow e^{-i2\pi f t_0}\ X(f)$$

- Modulation

$$x(t)e^{-i2\pi f_0 t} \Leftrightarrow X(f - f_0)$$

- Parseval's theorem

$$\int_{\Re} |x(t)|^2 = \int_{\Re} |X(f)|^2$$

The DFT is the sample Fourier transform. Therefore, it does not contain all frequencies forming an image, but only a set of samples that are large enough to perfectly describe the spatial domain image. The number of frequencies corresponds to the number of pixels

in the spatial domain image, i.e. the image in the spatial and Fourier domain are of the same size. For a square image size $N \times N$, the two-dimensional DFT is given by

$$F(k,\ell) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) e^{-i2\pi(\frac{ki+\ell j}{N})} \tag{2.4}$$

where $f(i,j)$ is the image in the spatial domain and the exponential term is the basis function corresponding to each point $F(k,\ell)$ in the Fourier space. Eq. 2.4 can be interpreted as: the value of each point $F(k,\ell)$ is obtained by multiplying the spatial image with the corresponding base function and summing up the result. The basis functions are sine and cosine waves with increasing frequencies, i.e. $F(0,0)$ represents the DC-component of the image which corresponds to the average brightness and $F(N-1, N-1)$ represents the highest frequency. In a similar manner, the Fourier image can be re-transformed to the spatial domain. The inverse Fourier transform is given by

$$f(i,j) = \frac{1}{N_2} \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} F(k,\ell) e^{i2\pi(\frac{ki+\ell j}{N})} \tag{2.5}$$

2.2 Fourier-Mellin Transform (FMT)

The Fourier-Mellin transform is a useful mathematical tool for image recognition or image onto image recognition and image database retrieval, because its resulting spectrum is invariant in rotation, translation and scaling (Lin et al., 2001; Swaminathan et al., 2006). Let f denote a function representing a gray-level image defined over a compact set of $\mathbb{R}^2$. The standard Fourier-Mellin transform of f is given by:

$$\forall (k,v) \in \mathbb{Z} \times \mathbb{R}, M_f(k,v) = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} f(r,\theta) r^{-iv} e^{-ik\theta} d\theta \frac{dr}{r}, \tag{2.6}$$

for $\forall (k,v) \in \mathbb{Z} \times \mathbb{R}$. $\mathbb{Z}^1$ denotes the additive group of integers. $\mathbb{R}$ denotes the additive group of the real line. f is assumed to be summable over $\mathbb{R}^* \times \mathbb{S}^+$ ($R^* \times$ denotes additive

group of integers and R^2) under the measure $d\theta\frac{dr}{r}$, i.e.

$$\int_0^\infty \int_0^{2\pi} |f(r,\theta)r^{-iv}e^{-ik\theta}|d\theta\frac{dr}{r} = \int_0^\infty \int_0^{2\pi} \frac{1}{r}f(r,\theta)d\theta dr < \infty, \tag{2.7}$$

since f is positive. Hence, the FMT could be divided into main three steps, which result in the invariance to rotation, scaling and translation attacks:

- *The Fourier Transform (FT)*: It converts the original image in spatial domain onto spectrum domain. The magnitude of Fourier transform itself is the translation invariant.

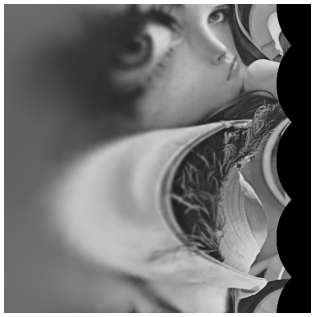
- *The Cartesian to Log-Polar Coordinates*: The conversion to log-polar coordinates converts the scale and rotation differences to vertical and horizontal offets that can be measured.

- *The Mellin Transform*: A second FT, called the Mellin transform (MT) gives a transform-space image that is invariant to rotation, scaling and translation.

(a) Original image

(b) Image in Log-polar transform

(c) Rotated image



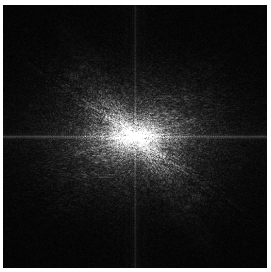
(d) Image in Log-polar transform

(e) Shifted image

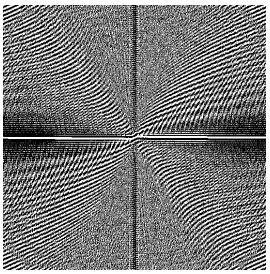
(f) Image in Log-polar coordinates

Figure 2.2: Example of Log-polar transform

(a) Original image

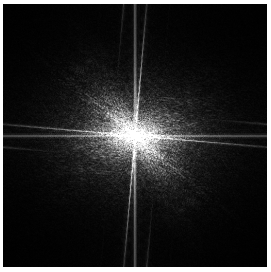


(b) FFT in Cartesian coordinates

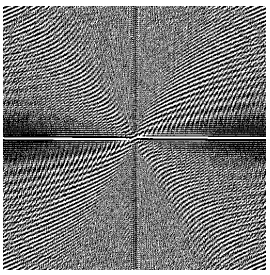


(c) FFT in Log-polar coordinates

(d) Rotated image

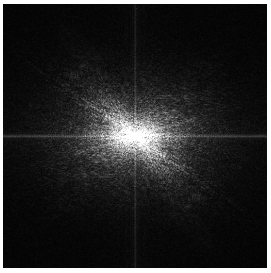


(e) FFT in Cartesian coordinates

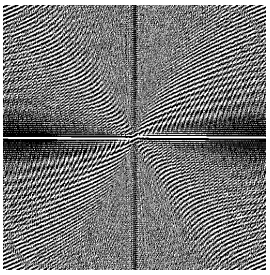


(f) FFT in Log-polar coordinates

(g) Shifted image



(h) FFT in Cartesian coordinates



(i) FFT in Log-polar coordinates

Figure 2.3: Example of Fourier-Mellin Transform

2.3 Discrete Cosine Transform (DCT)

An additional sinusoidal transform (i.e. transform with sinusoidal based functions) related to the DFT is the DCT. For an $N \times N$ image, the DCT is given by:

$$C(k,\ell) = \alpha(k,\ell) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos(\frac{(2i+1)k\pi}{2N}) \cos(\frac{(2j+1)\ell\pi}{2N}) \tag{2.8}$$

with

$$\alpha(k,\ell) = \begin{cases} \frac{1}{N} & for \quad k,\ell = 0 \\ \frac{2}{M^2} & for \quad k,\ell = 1,2,\ldots,N-1 \end{cases}$$

The main advantages of the DCT are that it gives a real output image and that it is a fast transform. A major use of the DCT is in image compression. Indeed, after performing a DCT it is possible to discard the coefficients representing high frequency components that the human eyes is not very sensitive to. Thus, the amount of data can be reduced, without seriously affecting the way an image appears to the human eyes.

In image compression, the DCT was established before the wavelet revolution and was the space coding adopted by the JPEG still image compression standard (Pennebaker and Mitchell, 1992). This transform is actually very close to the Karhunen-Loève Transform (KLT), which produces uncorrelated transforms coefficients of a Gaussian source (Jain et al., 1984).

(a) Original image (b) DCT domain

Figure 2.4: Discrete cosine transform of "Lena"

2.4 Discrete Wavelet Transform (DWT)

The wavelet decomposition is a mathematical tool allowing the study of signals and signal-generating processes characterised by a non stationary behaviour (Patrice, 1997). It accounts for the evolution in time of the frequency content of a signal. A signal $x(t)$ can often be better analysed, described, or processed if expressed as a linear decomposition by

$$x(t) = \sum_{j,k} a_{j,k} 2^{j/2} \psi(2^j t - k) \tag{2.9}$$

where the two-dimensional set of coefficients $a_{j,k}$ is called *discrete wavelet transform* (DWT) of $x(t)$. Note that the basis functions $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ are generated from a single function $\psi(t)$ called '*mother wavelet*' by changing two parameters j and k. The location of the wavelet moves in time or space, as the index k changes. This allows the expansion to explicitly represent the location of events in time or space and enables a representation of detail or resolution. A more precise way of indicating how the $a_{j,k}$'s are calculated can be written using the inner products as

$$x(t) = \sum_{j,k} \langle \psi_{j,k}(t), x(t) \rangle \psi_{j,k}(t) \tag{2.10}$$

2.4.1 Multiresolution Analysis

The multiresolution formulation of wavelet systems is designed to represent signals where a single event is decomposed into finer and finer details (Burrus et al., 1998). As described earlier for the wavelet, a set of scaling function is defined in terms of integer translates of the basic scaling function φ(t) by

$$\varphi_k = \varphi(t - k) \tag{2.11}$$

The subspace of $L^r\Re$ spanned by these functions is defined as

$$V_0 = \overline{Span_k(\varphi_k(t))} \tag{2.12}$$

A two-dimensional family of functions is generated from the basic scaling function by scaling and translation by

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k) \tag{2.13}$$

whose span over k is

$$V_j = \overline{Span_k(\varphi_k(t))} \tag{2.14}$$

This means that if $x(t) \in V_j$, then it can be expressed as

$$x(t) = \sum_k a_k \varphi(2^j t - k) \tag{2.15}$$

For $j > 0$, the span can be larger since $\varphi_{j,k}$ is narrower and is translated in smaller steps, The basic requirement of multiresolution analysis is

$$V_0 \subset V_1 \subset V_2 \ldots \subset L^2 \tag{2.16}$$

Hence, the spaces V_j satisfy a natural scaling condition

$$x(t) \in V_j \Leftrightarrow x(2t) \in V_{j+1} \tag{2.17}$$

The important features of a signal can be better described by also using a set of wavelet functions $\psi_{j,k}(t)$ that span the differences between the successive spaces V_j. Let us denote the orthogonal complement of V_j in V_{j+1} as W_j. It follows

$$V_1 = V_0 + W_0 \tag{2.18}$$

which extends to

$$V_n = V_0 + W_0 + W_1 + \ldots + W_{n-1} \tag{2.19}$$

Therefore, a signal $x(t) \in V_n$ can be expressed as

$$x(t) = \sum_k a_k \varphi(t - k) + \sum_{j=0}^{n-1} \sum_k d(j,k)\psi_{j,k}(t) \tag{2.20}$$

Eq. 2.23 represents a decomposition of $x(t)$ with n resolutions(or scales). The first summation gives a function that is a low resolution or coarse approximation of $x(t)$. For each increasing index j in the second summation, a higher or finer resolution function is added, which adds increasing detail. This is somewhat analogous to a Fourier series where the higher frequency terms contain the detail of the signal. From Eq. 2.16, it can be observed that if a function $\varphi(t)$ is in v_{j-1}, it is also in V_j, which is the space spanned by $\varphi(2^j t)$. This means $\varphi(2^{j-1}t)$ can be expressed in terms of a weighted sum of shifted $\varphi(2^j t)$

$$\varphi(2^{j-1}t) = \sum_n h(n) 2^{j/2} \varphi(2^j t - n) \tag{2.21}$$

Similarly, since $W_{j-1} \subset V_j, \psi(2^{j-1}t)$ can be expressed as

$$\varphi(2^{j-1}t) = \sum_n g(n) 2^{j/2} \varphi(2^j t - n) \tag{2.22}$$

Assume a signal x(t)$\in V_j$ which can therefore be written as

$$x(t) = \sum_k a_{j-1,k} 2^{(j-1)/2} \varphi(2^{j-1}t - k) + \sum_k d_{j-1,k} 2^{(j-1)/2} \psi(2^{j-1}t - k) \tag{2.23}$$

where

$$a_{j-1,k} = \langle x(t), 2^{(j-1)/2} \varphi(2^{j-1}t - k) \rangle = \int x(t) 2^{(j-2)/2} \varphi(2^{j-1}t - k) dt \tag{2.24}$$

and

$$d_{j-1,k} = \langle x(t), 2^{(j-1)/2} \varphi(2^{j-1}t - k) \rangle = \int x(t) 2^{(j-2)/2} \varphi(2^{j-1}t - k) dt \tag{2.25}$$

From Eq. 2.21 and 2.22 one can deduce

$$a_{j-1,k} = \sum_m h(m - 2k) a_{j,k} \tag{2.26}$$

and

$$d_{j-1,k} = \sum_m g(m - 2k) a_{j,k} \tag{2.27}$$

The last two equations represent a digital filtering process followed by a down-sampling (also called decimating) by a factor of 2. The down-sampler takes a signal $x(n)$ as an input and produces an output $y(n) = x(2n)$. These equations show that the scaling and wavelet coefficients at different levels of scale can be obtained by convolving the expansion coefficients at scale j by the time-reversed recursion coefficients $h(-n)$ and $g(-n)$ then down-sampling to give the expansion coefficients at the next level of j-1. In other words, the scale j coefficients are *filtered* by two FIR digital filters with coefficients $h(-n)$ and $g(-n)$. Subsequently, the down-sampling gives the next coarser scaling and wavelet coefficients. These structures implement Mallat's algorithm (Mallat, 1989) and have been developed in filter bank, quadrature mirror filters, conjugate filters, and perfect reconstruction filter bank in the literature (Smith and Barnwell, 1987; Vaidyanathan, 1987; Vetterli, 1987). Mallat, Daubechies, and others showed the relation of wavelet coefficient calculation and filter banks. The implementation of equations 2.26 and 2.27 is illustrated in Figure 2.5, where the down-pointing arrows denote a down-sampling by two and other boxes denote convolution by $h(-n)$ or $g(-n)$. This splitting, filtering, and decimation can be repeated on the scaling coefficients to give the two-scale structure in Figure 2.6

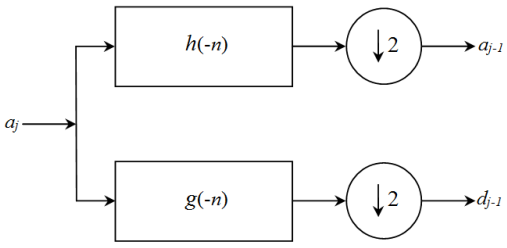


Figure 2.5: One-stage wavelet decomposition

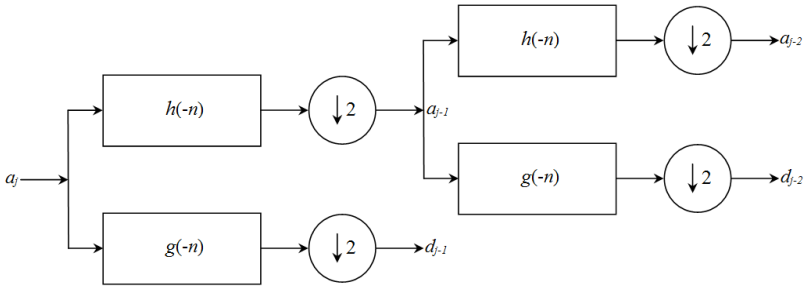


Figure 2.6: Two-stage wavelet decomposition

This splitting, filtering, and decimation can be repeated on the original fine scale coefficients and can be made from a combination of the scaling function and wavelet coefficients at a coarse resolution. This is derived by considering a signal in the j scaling function space $x(t) \in V_j$, which can be expressed as given by Eq. 2.23 or in terms of the scaling function at the same level j by

$$x(t) = \sum_k a_{j,k} 2^{j/2} \varphi(2^j - k) \tag{2.28}$$

Substituting Eq. 2.21 and 2.22 into 2.23 gives

$$x(t) = \sum_k a_{j-1,k} \sum_n h(n) 2^{j/2} \varphi(2^j t - 2k - n) + \sum_k d_{j-1,k} \sum_n g(n) 2^{j/2} \varphi(2^j t - 2k - n) \tag{2.29}$$

Because all of these functions are orthogonal, multiplying 2.28 and 2.29 by $\varphi(2^j t - k')$ and integrating evaluates the coefficient as

$$a_{j,k} = \sum_m a_{j-1,k} k(k - 2m) + \sum_m d_{j-1,k} g(k - 2m) \tag{2.30}$$

This final equation is actually evaluated by up-sampling the (j-1) scale coefficient sequence $a_{j-1,k}$, which means double its length by inserting zeros between each term, then convolving it with the scaling filter $h(n)$. The same procedure is performed to the (j-1) level wavelet coefficient sequence $d_{j-1,k}$ and the results are added to produce the j level scaling function coefficients $a_{j,k}$. This structure is illustrated in Figure 2.7. This process can be continued

to any level by combining the appropriate scale wavelet coefficients.

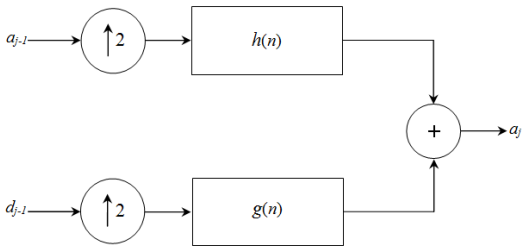


Figure 2.7: One-stage wavelet reconstruction

The resulting two-scale tree is show in Figure 2.8

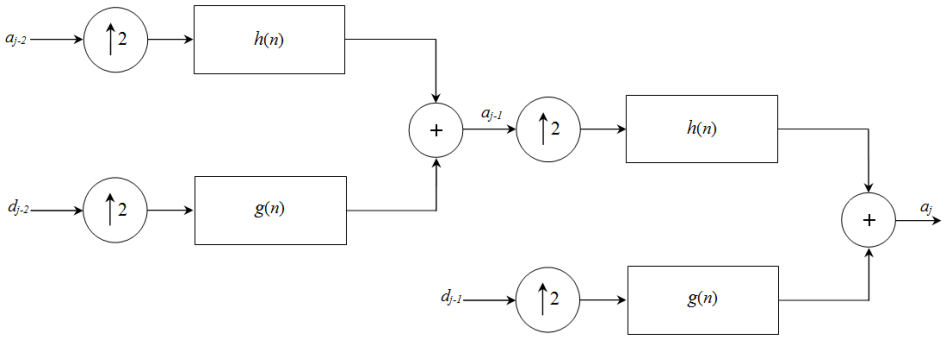


Figure 2.8: Two-stage wavelet reconstruction

2.4.2 Properties

In practical applications, wavelet bases can judiciously be chosen to fit the behaviour of the data to be analysed. An excellent choice of the wavelet bases can optimise coding and quantisation algorithms. Indeed, a wavelet basis that produces more coefficients with a magnitude closed to zero is preferred more in data compression, since these coefficients require less bits to encode. The most relevant criteria are the number of vanishing moments, the size of the support and regularity (Mallat, 1989; Villasenor et al., 1995).

The number of vanishing moments is related to the smoothness or differentiability of $\varphi(t)$ and $\psi(t)$. The size of the support measures the interval in time in which the wavelet takes non-zero values. Regularity is defined in terms of zeros of the frequency response function

of the scaling filter $h(n)$ thus, indicating how fast the Fourier transform magnitude drops off, as the frequency progresses to infinity. This is particularly related to the frequency localisation of the decomposed signal.
The size of the wavelet support increases with the number of vanishing moments. The wavelet regularity is important to reduce the artefacts. The choice of an optimal wavelet in image compression is thus the result of a trade-off between the number of vanishing moments and artefacts (Antonini et al., 1992). Some useful properties of the wavelet transform can be summarised as follows:

1. They can represent smooth functions.
2. They can represent singularities.
3. The basis functions are local. This makes most coefficient-based algorithms naturally adaptive to inhomogeneities in the function.
4. They have the unconditional basis property for a variety of function classes implying that if one does not know much about a signal (for instance, a signal with a non stationary behaviour), the wavelet basis is usually a reasonable choice.
5. They are computationally inexpensive with a complexity $O(N)$ compared to a Fourier transform, which is $N\log(N)$ or an arbitrary linear transform which is $O(N^2)$.

2.4.3 2-D wavelet transform

For 2-D data such as images, the most commonly used algorithm for wavelet decomposition uses separable one-dimensional wavelets and scaling functions (Mallat, 1989).
This kind of two-dimensional DWT leads to a decomposition of approximation coefficients at level j in in four components: the approximation at level $j-1(a_{j-1})$, and the details in three orientations (horizontal $d_{j-1}^{(h)}$, vertical $d_{j-1}^{(v)}$, and diagonal $d_{j-1}^{(d)}$. Figure 2.9 describes the basic decomposed steps for images.
An example of a one-stage decomposed image of "Lena" is illustrated by Figure 2.10. In a similar way, the reverse process can be used to obtain the original 2-D signal.

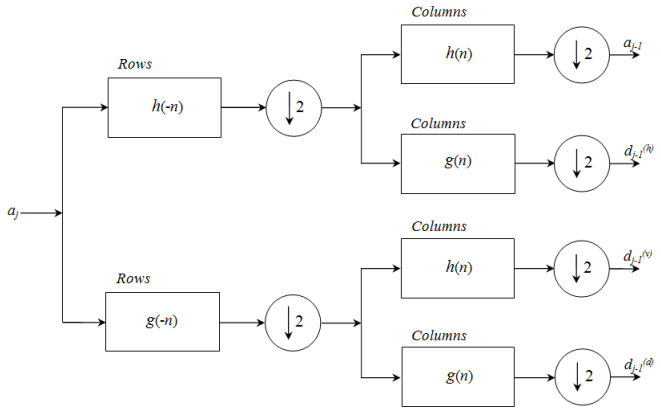


Figure 2.9: One-stage 2-D wavelet decomposition

(a) Original image (b) Decomposed image

Figure 2.10: One-stage 2-D wavelet decomposition of "Lena"

2.5 Conclusion

In this chapter, the basis and characteristics of 4 popular discrete transforms, namely Discrete Fourier Transform (DFT), Fourier-Mellin Transform (FMT), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT), are reviewed. This is due to the fact that these 4 techniques are currently very popular, within image processing. Such transforms are the most challenging part of the image hashing in the feature extraction stage, in order to extract image features that are invariant to content-preserving image processing operations.

Chapter 3

Proposed perceptual image hashing in the DWT domain with non-negative matrix factorisation (NMF)

Recently, the discrete wavelet transform (DWT) has been successfully used in several image hashing algorithms and was reported to outperform previous techniques (Antonini et al., 1992; Venkatesan et al., 2000; Mihçak and Venkatesan, 2002; Fawad and Siyal, 2006; Hu and Niu, 2010; Hassan et al., 2012), because it characterises the image content both in the spatial and the frequency domains. For instance, DWT based robust image hashing was explored in Hu and Niu (2010). It has been shown to be highly robust to image processing operations motivating other solutions in this direction. Monga and Mihçak (2007) presented a new image hashing algorithm by using a dimension reduction technique, called non-negative matrix factorisation (NMF) (Lee and Seung, 2001). The advantage of NMF hashing is the structure of basis resulting from its non-negative basis vectors, which leads to a parts-based representation. Based on the results by Monga and Mihçak (2007), the NMF hashing possesses a good robustness under perceptually insignificant attacks.

Inspired by the potential of DWT and NMF for image hashing, a robust and secure image hashing based on a pseudo-random sub-images selection in the DWT domain and NMF is presented (Prungsinchai et al., 2012). The basic idea of the proposed image hashing technique is to first transform the sub-images into DWT domain. The low frequency sub-band coefficients are used, and then NMF is applied to obtain robust image features. The final step of the image hashing algorithm is to generate the binary image hash. The objective of this method of analysis is to test the robustness of the image hashes generated from the coefficients of the LL sub-band and in addition, the security of the hashing system. To show the advantages of the proposed hashing technique, a number of hashing algorithms have also been applied on the same images for a fair comparison including SVD-based image hashing (Kozat et al., 2004), feature points-based image hashing (Monga et al., 2005), and DWT-NMF-based image hashing (Hassan et al., 2012) which was reported to significantly outperform other existing hashing approaches. Our preliminary experimental results (Prungsinchai et al., 2012) showed that the proposed image hashing technique offers better identification performance under various attacks such as JPEG lossy compression, media filtering, additive noise etc. Section 3.1 provides a theoretical background. The proposed framework for the perceptual image hashing scheme in the DWT domain is shown in section 3.2. Section 3.3 presents an identification and evaluation measure of image hashing. Section 3.4 presents experimental results and analysis: robustness testing, robustness versus discriminability testing and unpredictability testing. Finally, section 3.5 summarises the key ideas introduced in the chapter.

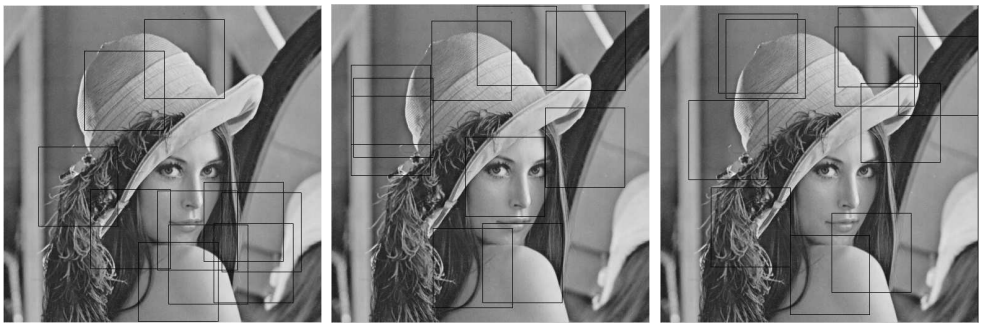
3.1 Theoretical background

3.1.1 Pseudo-randomly partition

The idea of pseudo-randomly selecting sub-regions or sub-images form an original input image, as shown in Figure 3.1, has been used by Venkatesan (Venkatesan et al., 2000), Monga (Monga and Mihçak, 2007), and Jie (Jie, 2013). Venkatesan's method (Venkate-

san et al., 2000) and Mihçak's method (Mihçak and Venkatesan, 2002) introduced image hashing security by producing features from pseudo-randomly rectangles from which features were generated. Monga's method (Monga and Mihçak, 2007) and Hernandez's method (Hernandez et al., 2011) employed a pseudo-randomly partitioning of the image to introduce unpredictability in the hash values. Using pseudo-randomly partition algorithms is desirable for security and also adds unpredictability of the hash value, which in turn motivated us to use this technique to enhance the security of the hashing system. The pseudo-randomly sub-images technique is described as follows:

- ***Pseudo-randomly coordinates***: Pseudo-randomly select a P set of coordinates (x,y) depending on the secret key and acquire $P(x_i, y_i)$, for $1 \leq \mathrm{i} \leq P$.
- ***Generate region***: Each $P(x_i, y_i)$ is used to generate a square region as the based point top-left corner of square region with size N-by-N , which is called sub-image.



(a) Secret key K_1 (b) Secret key K_2 (c) Secret key K_3

Figure 3.1: Example of pseudo-random generating sub-images of "Lena" with different secret keys with size N-by-N

3.1.2 Non-negative Matrix Factorisation (NMF)

The NMF approach is a popular dimensionality reduction technique that has many useful applications and has been successfully applied to a variety of tasks in various fields e.g. computer vision, text mining, etc. NMF is applied to high dimensional data and it provides a low rank approximation form (Lee and Seung, 1999, 2001; Li and Fukui, 2007; Monga

and Mihçak, 2007; Korattikara et al., 2011). Given an image V with size $n \times m$, V can be approximately factorised into the product of two non-negative matrices W with size $n \times r$ and H with size $r \times m$ (see Figure 3.2), where r is the rank of decomposition. In practice, r is usually smaller than n and m, so that the non-negative matrices of W and H are smaller than the original matrix V. The matrix W with size $n \times r$ contains the NMF basic vector, and the weight matrix with the H with size $r \times m$ contains the associated coefficients.

$$V \approx WH \tag{3.1}$$

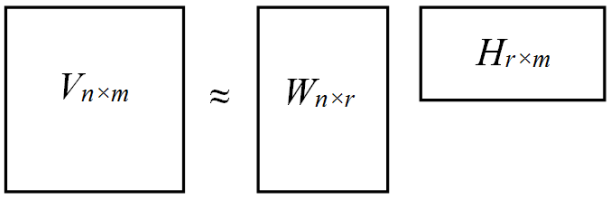


Figure 3.2: Illustration of the NMF approximation

Cost functions

To measure the quality of the approximate factorisation $V \approx WH$, a cost function can be constructed by using a measure of distance between two non-negative matrices W and H. The two popular cost functions are the classical Euclidean distance or Frobenius norm, given by (Pentti, 1997):

$$\| V - WH \|^2 = \sum_{i,j} (V_{ij} - (WH)_{ij})^2 \tag{3.2}$$

Another measure commonly used in practice is V from (WH):

$$D(V\|WH) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (V_{ij} \log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}) \tag{3.3}$$

The above measure is known as the generalised Kullback-Leibler (KL) divergence. It

reduces to the standard Kullback-Leibler divergence, or relative entropy, when $\sum_{ij} V_{ij} = \sum_{ij}[WH]_{ij} = 1$, so that the matrices V and WH can be regarded as normalized probability distributions. The lack of convexity of the aforementioned costs in both factors W and H means it is unrealistic to expect a close from solution. In this work, the multiplicative update rules were employed (Pentti, 1997) to find W and H as follows:
The multiplicative update rules were employed to find W and H as given by Pentti (1997); Lee and Seung (2001)

Theorem 1 *The Frobenius or Euclidean distance* $\|V - WH\|$ *does not increase the update rules*:

$$H_{lj} \leftarrow H_{lj} \frac{(W^T V)_{lj}}{(W^T W H)_{lj}} \qquad\qquad W_{il} \leftarrow W_{il} \frac{(V H^T)_{il}}{(W H H^T)_{il}} \tag{3.4}$$

Further, $\| V - WH \|$ is invariant under these updates if and only if W and H are at a stationary point of the distance.

Theorem 2, *The divergence* $D(\|V - WH\|)$*is non-increasing under update rules*:

$$H_{lj} \leftarrow H_{lj} \frac{\sum_{i=0}^{m-1} W_{il} V_{ij} / (WH)_{ij}}{\sum_{i=0}^{m-1} W_{il}} \qquad W_{il} \leftarrow W_{il} \frac{\sum_{j=0}^{n-1} H_{lj} V_{ij} / (WH)_{ij}}{\sum_{j=0}^{n-1} H_{lj}} \tag{3.5}$$

where $i = 0, 2, \ldots, m\text{-}1$, $j = 0, 2, \ldots, n\text{-}1$, $l = 0, 2, \ldots, r\text{-}1$. Furthermore, $D(V\|WH)$ is invariant under these updates, if and only if W and H are at a stationary point of the divergence. Proof of these theorems can be found in Pentti (1997). It can be observed that the updates are multiplicative. It is also straightforward to see that the multiplicative factor is agreement when $V = WH$, so that the perfect reconstruction is necessarily a fixed point of the update rules.

3.2 Proposed perceptual image hashing in DWT domain with non-negative matrix factorisation (NMF)

The proposed perceptual image hashing scheme is illustrated in Figure 3.3. The image hashing system operating in the DWT domain consists of four steps:

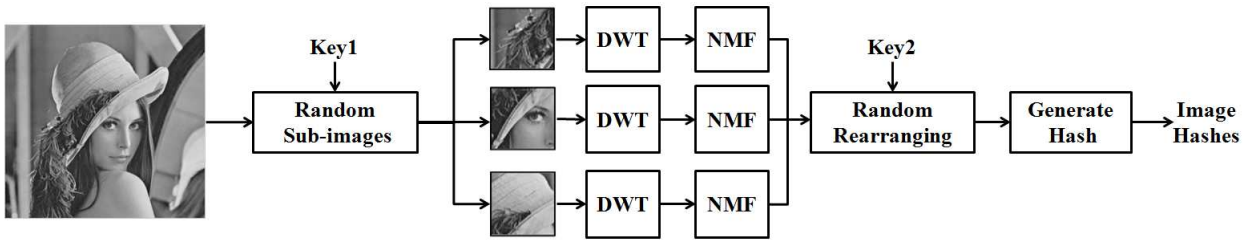


Figure 3.3: The proposed DWT-based image hashing scheme

- ***The pseudo-random sub-images step***: Given an input image I with size M-by-M. Pseudo-randomly based on secret key 1 to generate a set of coordinate pairs $P(x_i, y_i)$. Each coordinate pair $P(x_i, y_i)$ corresponds to a sub-region A_i with size N-by-N, which is referred here to as sub-image.

- ***The feature extraction step***: 3-levels of the wavelet decomposition are applied to each of the sub-images. A 3-level wavelet decomposition is shown in Figure 3.4, where L represents low pass filtering and H stands for a hight-pass filtering. The information of the approximation sub-band is a coarse version of the original image and contains all the perceptual information of the image. Subsequently, the LL sub-band is taken to the next step.

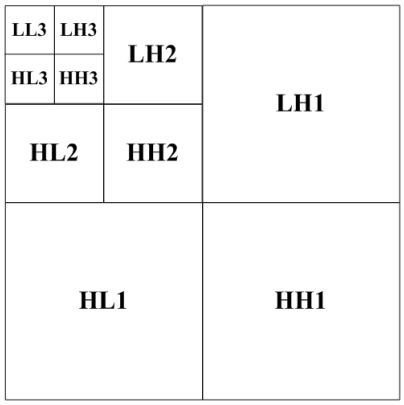


Figure 3.4: Wavelet decomposition with 3 levels

- ***The dimension reduction step***: In view of 3.1.2, NMF is applied to each sub-image R_i to yield the coefficient matrix H^m with size $r \times m$ denoted as C_i. All the coefficient matrix C_i are used as features which form a vector V^c.

- ***The hash computation step***: The vector V^c is interleaved based on secret key 2 resulting in $V^{c'}$. Finally, the interleaved vector is used to generate the binary hash values V^h. A binary string is obtained by considering the difference between neighbouring coefficients as follows:

$$V_i^h = \begin{cases} 0, V_i^{c'} \leq V_{i+1}^{c'} \\ 1, V_i^{c'} > V_{i+1}^{c'} \end{cases}$$

3.3 Identification and similarity measure

Perceptual robustness is one critical criterion used to evaluate the performance of the proposed image hashing algorithm. Ideally, a similar image $\hat{I}$ of the original image I under content-preserving distortions (CPOs) should have similar hashes, while two perceptually different images I and J should have different image hashes. In this work, the evaluation of this process is conducted in two aspects: identification accuracy and Receiver Operating Characteristics (ROC) analysis.

3.3.1 Identification process

The identification accuracy is defined as the fraction of the distorted image copies that are correctly classified as versions of the original image. $h_1(i)$ is the binary hash of the original image and $h_2(i)$ is the binary hash of another image. The normalised Hamming distance could be used as a criterion to measure the similarity between two binary image hashes H_1 and H_2, and can be defined as:

$$NHD(H_1, H_2) = \frac{1}{n} \sum_{i=1}^{n} |h_1(i) \oplus h_2(i)| \tag{3.6}$$

where n is the hash length. Recall from section 1.2, we consider the following requirements:

- normalised Hamming distance between the original image I and a similar version $\hat{I}$ should be close to 0.
- normalised Hamming distance between the original image I and a different image J should be close to 1.

3.3.2 Receiver Operating Characteristics analysis

The ROC curve is used, in which it depicts the relative tradeoff between TPR (benefit) and FPR (cost) of the identification process. It is used to compare the performance of different image hashing techniques. To obtain the ROC curve and to analyse the image hashing algorithms, the $TPR(\tau)$ and $FPR(\tau)$ are defined as:

$$TPR(\tau) = Probability(D(H(I, K), H(\hat{I}, K)) < \tau) \tag{3.7}$$

$$FPR(\tau) = Probability(D(H(I, K), H(J, K)) < \tau) \tag{3.8}$$

where τ is the identification threshold. I and J are two perceptually different or distinct image, and the image $\hat{I}$ is a version derived from the original image I. ROC curves were generated by varying the threshold τ from the minimum to the maximum value of all

distances. TPR against FPR were plotted in ROC curves which suggest that the best possible performance should correspond to a point in the top left corner (coordinate 0,1) of the ROC space.

3.3.3 Database and content-preserving operations

To evaluate the performance of the proposed image hashing algorithm, the dataset was constructed with 120 original gray scale natural images. This included around 90 classic benchmark images such as Lena, Baboon, Peppers, etc., and a variety of scenery and human pictures, which were mainly selected from three group of categories in the content-based image retrieval database of the University of Granada (CVG-URG, 2007). For each original image with size 512×512, six classes of content-preserving operations (CPOs) including JPEG lossy compression, median filtering, AWGN, rotation, translation and histogram equalisation were performed with various parameters on each original image as listed in Table 3.1. For this experiment, all the operations were implemented using MATLAB. For image rotation attack, a black frame around the image was added by MATLAB, however some parts of the image were cut to keep their size the same as the new image (see an example in Figure 1.2 (f) Rotation).

Table 3.1: Content-preserving operations with various parameters

Manipulation type	**Parameters**
Image processing operations	
JPEG lossy compression	quality factor $QF = 10\sim90$
Additive White Gaussian Noise (AWGN)	standard deviation $\sigma = 20\sim35$
Median filtering	window size $3\sim9$
Histogram equalization	/
Geometric distortions	
Rotation	degree $3^\circ\sim9^\circ$
Translation	window size $5\sim20$

3.4 Experimental results with the proposed image hashing technique in DWT domain

Following the discussion given in section 3.2, the proposed image hashing algorithm will be evaluated in two aspects. The first one is in relation to its perceptual robustness against content-preserving operations (CPOs), which is important for content identifications issues. It is desired that perceptually identical images under distortions would have similar image hashes. The second one is the discriminative capability. It is believed that perceptually different images would have different image hashes. To assess the proposed image hashing algorithm with different parameters given by $P = 16$ and 20, $N = 64$, 128 and 256 and $r = 1$, as shown in Table 3.2. The hash length is dependent on the length of the number of sub-images, size of sub-images and rank (r) of the decomposition as shown in Table 3.3. The same secret key is used to generated image hash for different images.

Table 3.2: Parameter setting in the proposed image hashing algorithm

Parameter	**Value**
Number of the sub-images	$\boldsymbol{P}$ = 16 and 20
Size of sub-images	$\boldsymbol{N}$ = 64,128 and 256
Rank of NMF	$\boldsymbol{r}$ = 1
Level of HAAR wavelet	$\boldsymbol{L}$ = 3

Table 3.3: Length of hash

Block size	**Number of sub-images**	
$\boldsymbol{N}$	$\boldsymbol{P = 16}$	$\boldsymbol{P = 20}$
64	128 (binary)	160 (binary)
128	256 (binary)	320 (binary)
256	512 (binary)	640 (binary)

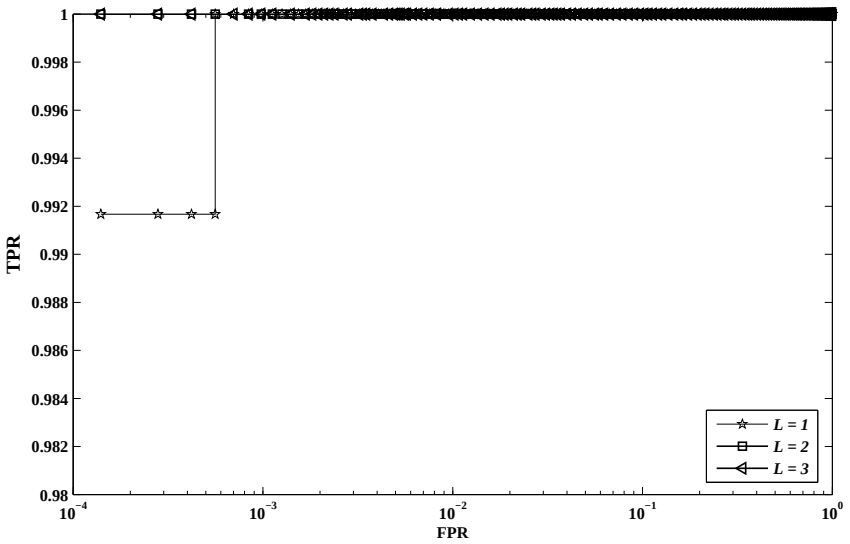
3.4.1 Robustness testing

Robustness implies that image hashing functions should be robust in content-preserving operations (CPOs) that include JPEG lossy compression, median filter, AWGN, histogram equalisation, rotation and translation, as the underlying content is perceptually identical to the HVS. The normalised Hamming distance (NHD) between the image hashes of the original image and the image hash of the operated image should be close to 0. To demonstrate robustness, we first investigated the image hashing technique with regard to the use of the size of sub-image and number of sub-images under different attacks for "Lena" image and the experimental results are plotted in Figures 3.6 to 3.11. It is clear that the proposed image hashing consistently yields a higher robust/identification accuracy under different types of tested content-preserving operations. The JPEG lossy compression (with variance level: $90 \backsim 10$) is shown in Figure 3.6, median filtering (with variance level: $3 \backsim 5$) in Figure 3.7, and AWGN (with various level: $20 \backsim 40$) in Figure 3.8, it can be perceived that for theses attacks the normalised Hamming distance (NHD) were lower than 0.03. Figure 3.9 shows the histogram equalisation, the normalised Hamming distance (NHD) was closer to 0.1. For the rotation (with various level: $3 \backsim 5$) as shown in Figure 3.10 and translation (with various level: $5 \backsim 15$) as shown in Figure 3.11, the normalised Hamming distance (NHD) from both rotation and translation attacks were lower than 0.5.

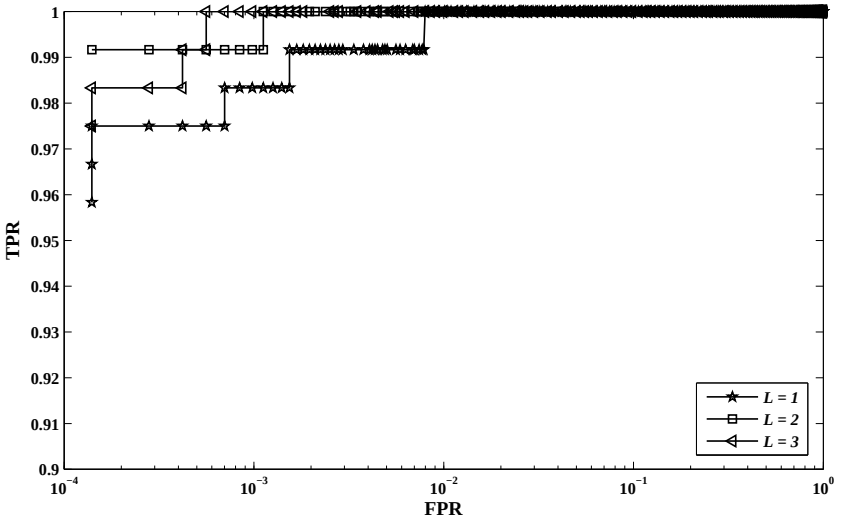
We also test the effect of the number of decompositions and the rank for NMF as shown in Figure 3.5, where the receiver operating characteristics (ROC) graph is exploited to visualise the classification performance with respect to the robustness and the discriminative capability under different parameters. In the experiments, the used rank for NMF are 1 and 3, i.e., $r = 1$ and $r = 3$. For each rank, three number of decompositions are consider, i.e., $L = 1$, $L = 2$ and $L = 3$. The number of the sub-images and size of sub-images are $P = 20$ and $N = 128$, respectively. The test images used are discussed in subsection 3.3.3.

It is observed that, for fixed rank, the whole image hashing performance improves when the number of decompositions increases. For a fixed number of decompositions, a bigger rank deteriorates the image hashing performance. This is because large rank provide more information in the hash about the image enhancing the discriminative capability, but not

necessarily preserving the perceptual robustness. Generally, the image hashing performance is related to hash length. A short image hash length could offer robust hashes but these may fail to discriminate different images. As image hash length increases, discriminative capability is strengthened while perceptual robustness decreases. An image hashing is a compact representation, meaning that the image hash length is expected to be as short as possible. Therefore, a trade-off is needed in choosing the image hash length, i.e., the values of L and r. From the experiments, we found that the choice of $P = 20$, the size of blocks $N = 128$, rank for NMF $(r) = 1$ and number of decompositions $(L) = 3$ offered the best performance and can reach a desirable compromise between the robustness and the discriminability. In the rest of this chapter, the parameters $P = 20$, $N = 128$, $r = 1$ and $L = 3$ will be adopted.

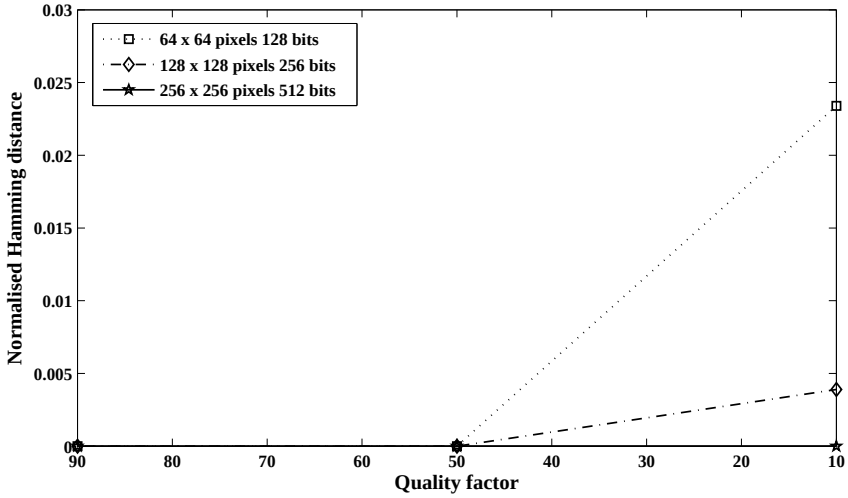


(a) $r = 1$

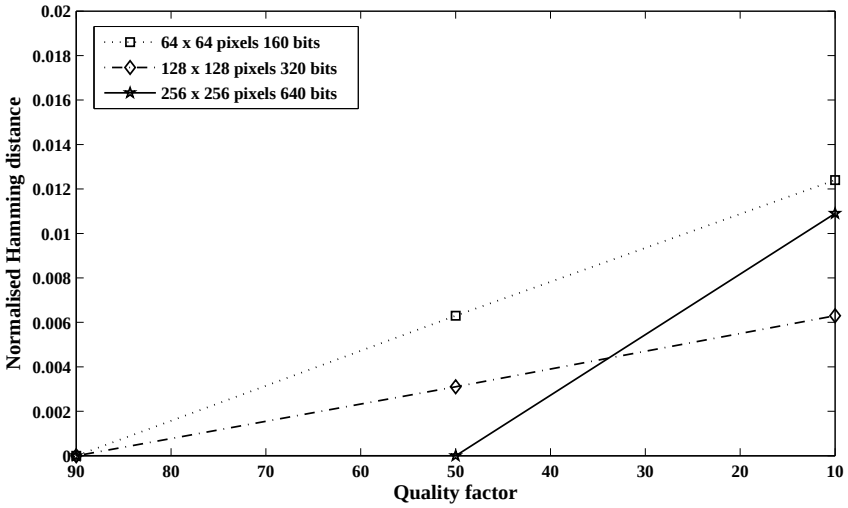


(b) $r = 3$

Figure 3.5: ROC curves under different parameters

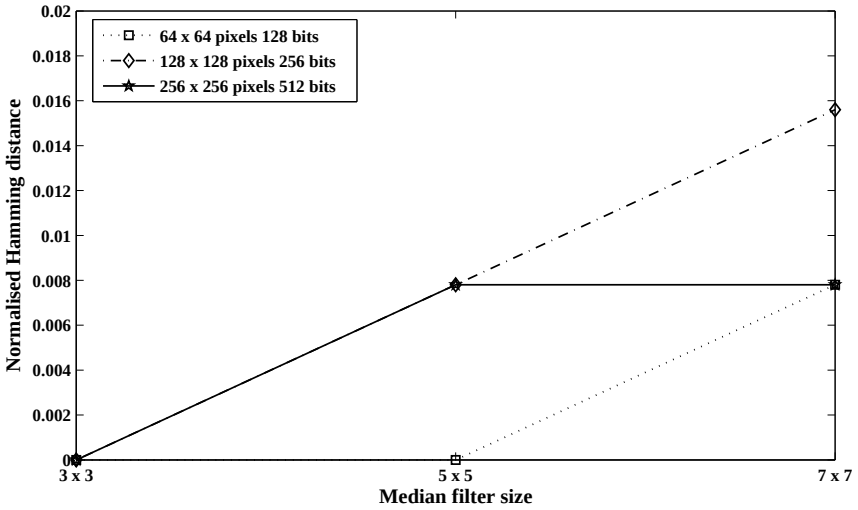


(a) 16 sub-images

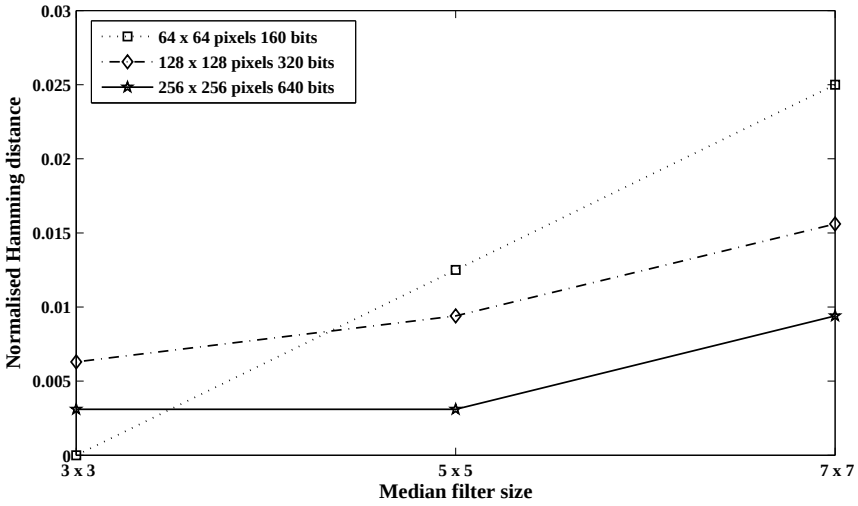


(b) 20 sub-images

Figure 3.6: Robustness evaluation of the proposed technique for "Lena" image with different sub-image sizes and a number of sub-images under JPEG lossy compression attack

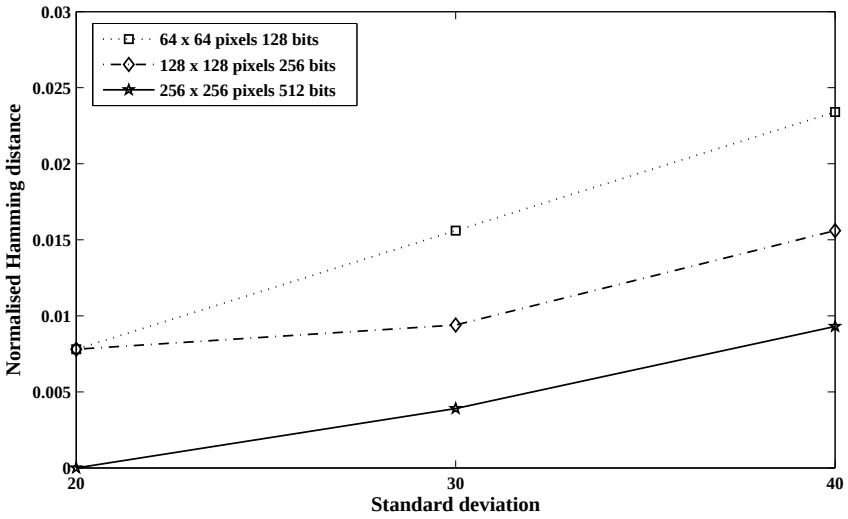


(a) 16 sub-images

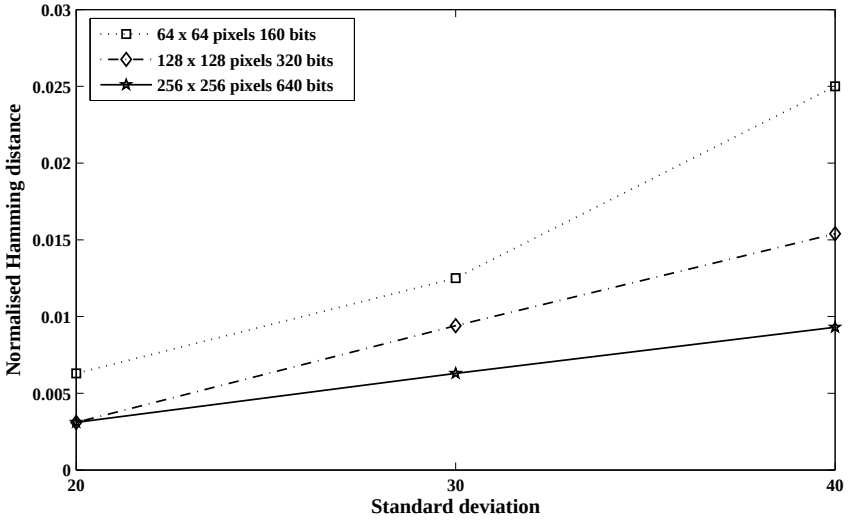


(b) 20 sub-images

Figure 3.7: Robustness evaluation of the proposed technique for "Lena" image with different sub-image sizes and a number of sub-images under a median filter attack

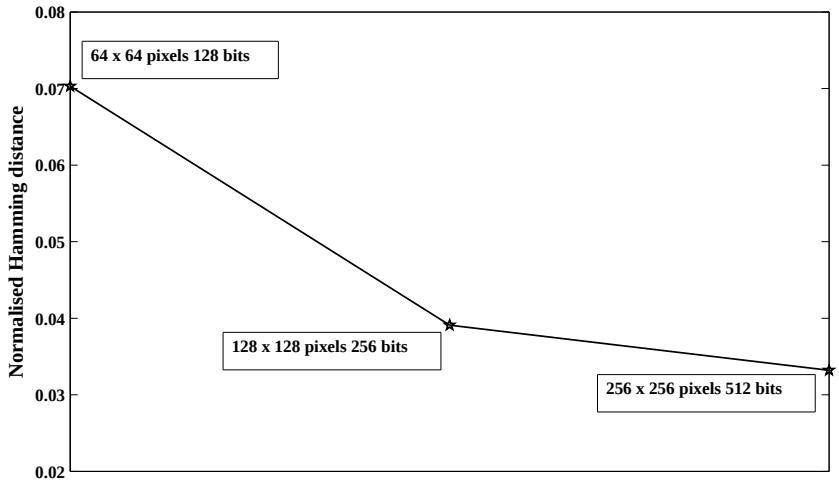


(a) 16 sub-images

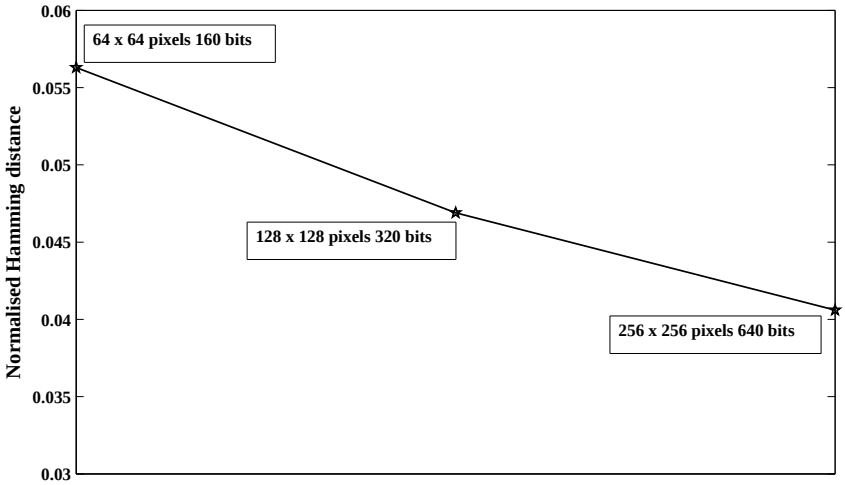


(b) 20 sub-images

Figure 3.8: Robustness evaluation of the proposed technique for "Lena" image with different sub-image sizes and a number of sub-images under AWGN attack

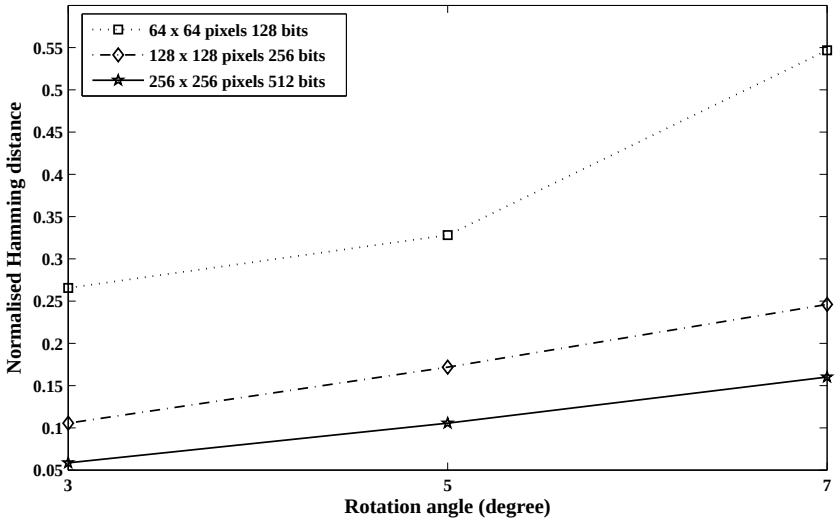


(a) 16 sub-images

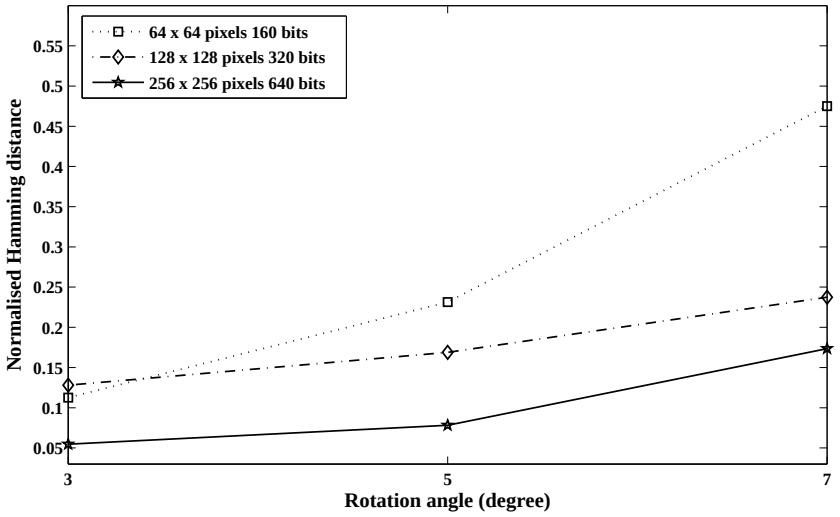


(b) 20 sub-images

Figure 3.9: Robustness evaluation of the proposed technique for "Lena" image with different sub-image sizes and a number of sub-images under a histogram equalisation attack

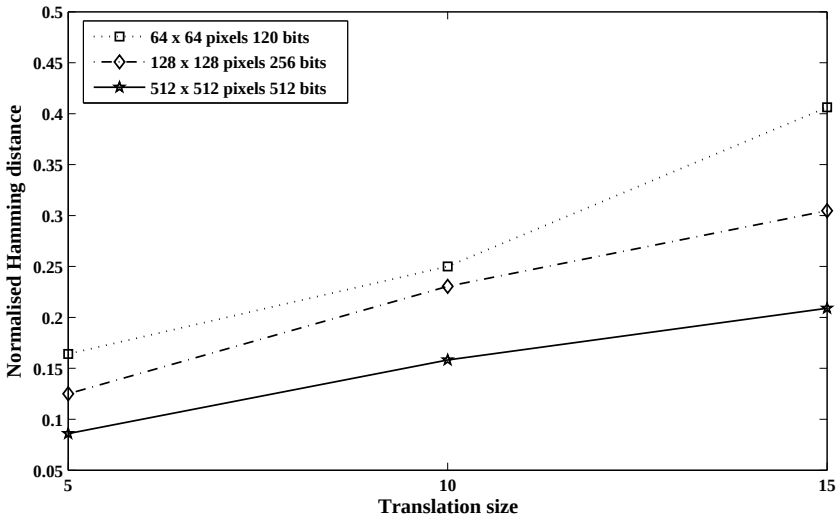


(a) 16 sub-images

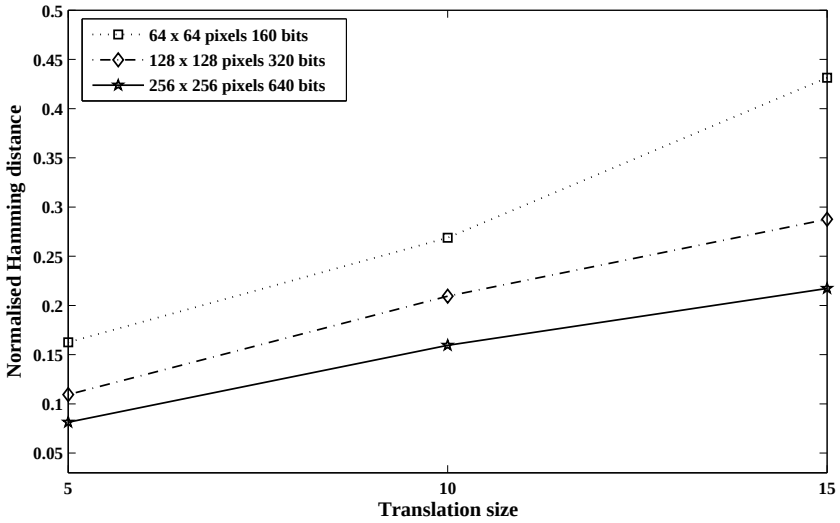


(b) 20 sub-images

Figure 3.10: Robustness evaluation of the proposed technique for "Lena" image with different sub-image sizes and a number of sub-images under a rotation attack



(a) 16 sub-images



(b) 20 sub-images

Figure 3.11: Robustness evaluation of the proposed technique for "Lena" image with different sub-image sizes and a number of sub-images under a translation attack

3.4.2 Robustness versus discriminability

A small normalised Hamming distance (NHD) between hash vectors extracted from the original and its attacked images in section 3.4.1 does not necessarily mean that the image

hashing system is reliable unless it could distinguish efficiently between visually different images through different hashes. ROC curves are exploited to visualise the classification performance and this is an effective way to compare different image hashing techniques with respect to the robustness and the discriminative capability. Let d_i be the distances between the hashes extracted from the original and the attacked images, d_i (i = 1, 2,..., s). q is the distance between visually different images $\hat{d}_i$ (i = 1, 2,..., q). ξ is the number of different attacks used, which are listed in Table 3.1. The ROC curves were generated by varying the threshold value τ from minimum value to the maximum value of distance and record probabilities ($d < \tau$) referred here as TPR and probabilities ($\hat{d} < \tau$) referred here as FPR. Firstly, the calculated distances between the original and attacked images were obtained ($s \times \xi$) and secondly, the computed distances between each pair of different images were calculated and obtained with $q = 7140$ values for 120 images. It is worth mentioning that the distance was used to compute TPR and FPR depending on the nature of the hash. Indeed, since the DWT-NMF-based image hashing technique extracted binary hashes, the normalised Hamming distance (NHD) is used to measure the similarity between two binary hash vectors. The Euclidian Distance is used for real valued hashes extracted via the SVD-based image hashing technique. Finally, the Hausdorff distance was used for hashes corresponding to the coordinates of feature points extracted via the feature points-based hashing technique. The parameters and hash length obtained from each hashing technique are listed in Table 3.4.

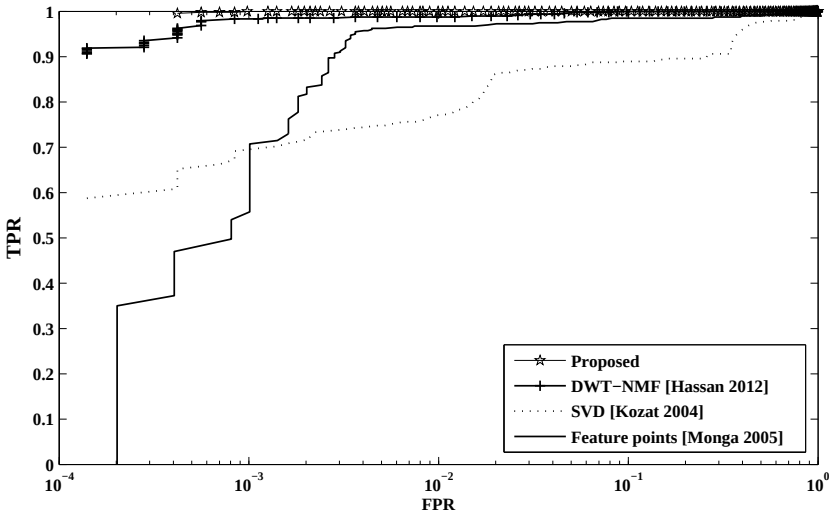
Table 3.4: Parameters used in the implementation and hash length

Hashing technique	**Parameters**	**Hash length**
Proposed	$P = 20$, $m = 128$, $r = 1$, and $L = 3$	320 (binary)
SVD	$p = 50$, $m = 256$, $d = 50$, and $r = 8$	800 (real values)
Feature points	64×2	128 (coordinates)
DWT-NMF	wavelet 3 levels, $m = n = 64$, and $r = 5$	320 (binary)

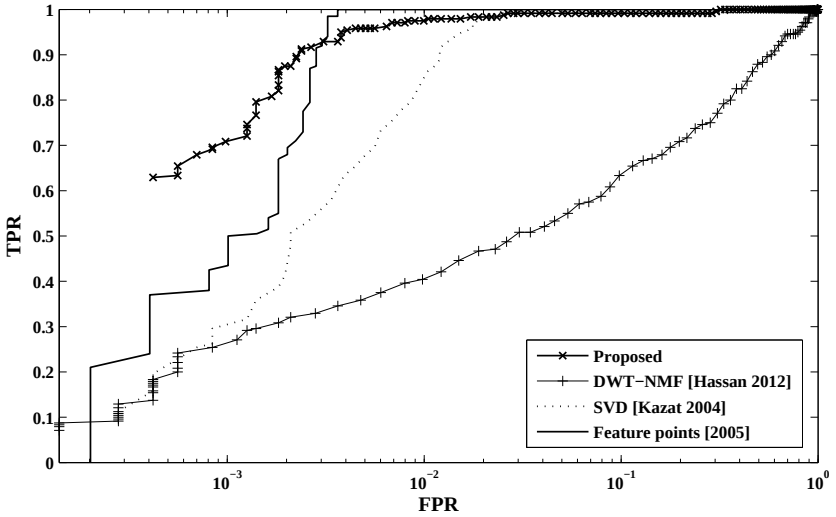
A statistical comparison of different image hashing techniques was investigated, by studying the corresponding ROC curves. The ROC curves provide a trade-off between the true

identification and misclassification. The overall ROC curves were generated for all types of test content-preserving operations (CPOs) when applying different hashing techniques are illustrated in Figure 3.12. Figure 3.12 (a), shows that the proposed image hashing technique achieved the best robustness and significantly outperforms related techniques under image precessing operations, closely followed by that the DWT-NMF-based image hashing technique. This is because the statistics of sub-images computed in the LL sub-bands are invariant. The feature point's image hashing technique presented a slightly lower performance under image processing attacks because their robustness against image processing operations, especially additive noising and blurring is limited. The SVD-based image hashing technique showed high sensitivity under image processing attacks, while it performs well under geometric attacks. Figure 3.12 (b), shows that the feature points image hashing technique offers the best performance under geometric attacks, closely followed by the proposed image hashing technique, within a range of value for TPR > 0.927 and FPR $> 10^{-2.518}$. Beyond this range, the proposed technique perform better. Observe that the DWT-NMF-based image hashing technique provides poor performance under geometric attacks.

To test the robustness to each type of content-preserving operations (CPOs), an ROC curve was also generated for each operation. The ROC curves corresponding to the six attacks (i.e. JPEG lossy compression, median filter, AWGN, histogram equivocation, rotation and translation) are shown in Figures 3.13 to 3.15. Once again, the ROC curves in Figures 3.13 to 3.14 reinforce the observation that the proposed image hashing technique significantly outperforms other state-of-the art techniques of DWT-NMF-based image hashing, SVD-based image hashing and feature points-based image hashing. However, the results shown in Figure 3.15 the feature points-based image hashing performs slightly better than the proposed technique under rotation and translation attacks in the corresponding range of TPR > 0.883, FPR $> 10^{-2.492}$ and TPR > 0.914, FPR $> 10^{-2.996}$, respectively.

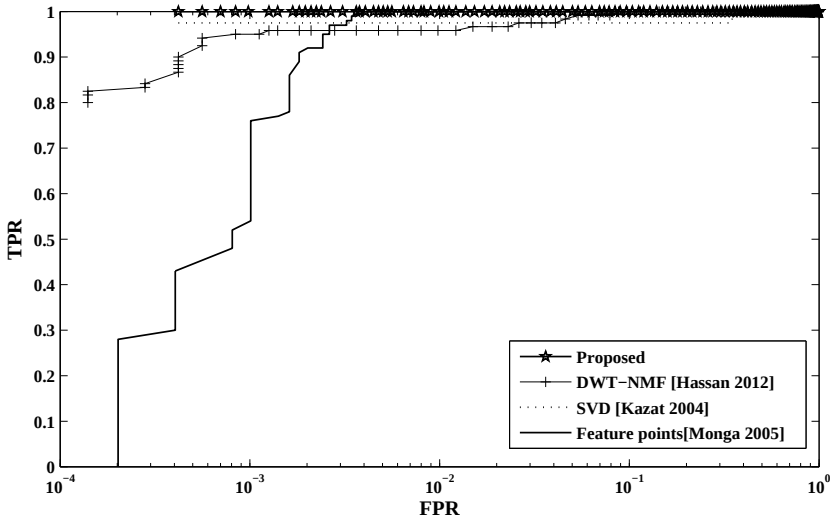


(a) Image processing operations

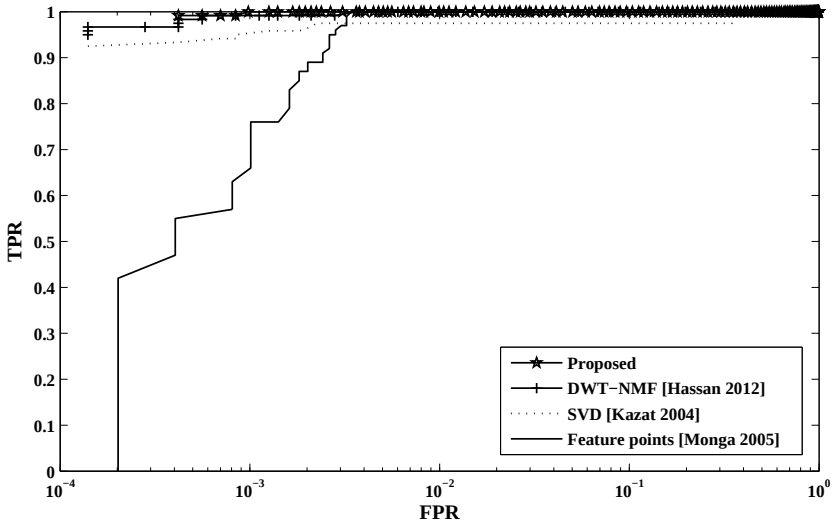


(b) Geometric attacks

Figure 3.12: The overall ROC curves for all types of test manipulations when applying different hashing techniques, (a) Image processing operations (b) Geometric attacks

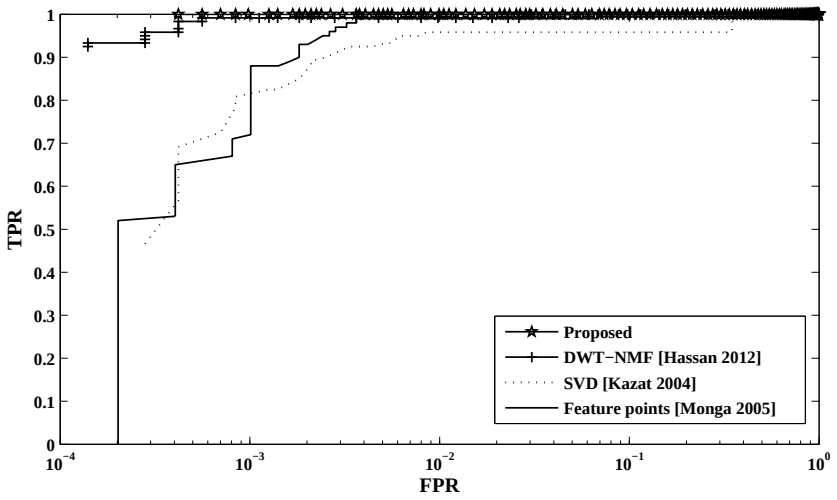


(a) JPEG lossy compression

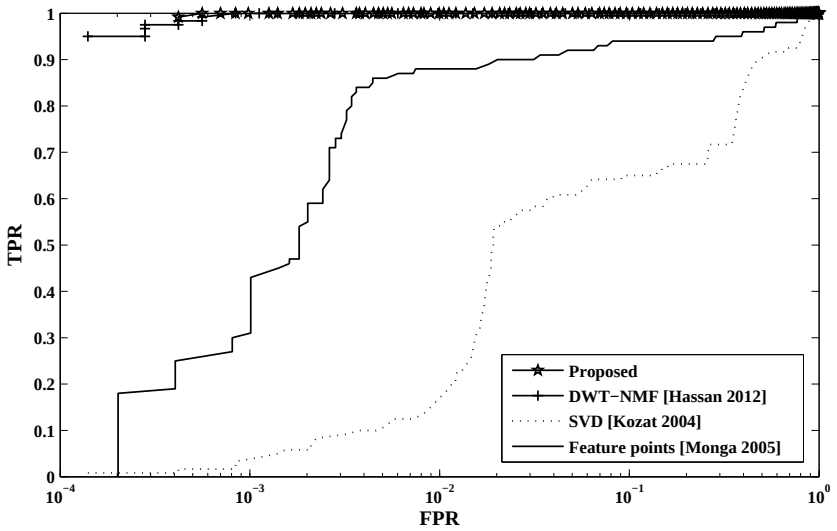


(b) Median filtering

Figure 3.13: ROC curves for each type of manipulations when applying into different image hashing techniques, (a) JPEG lossy compression (b) Median filtering

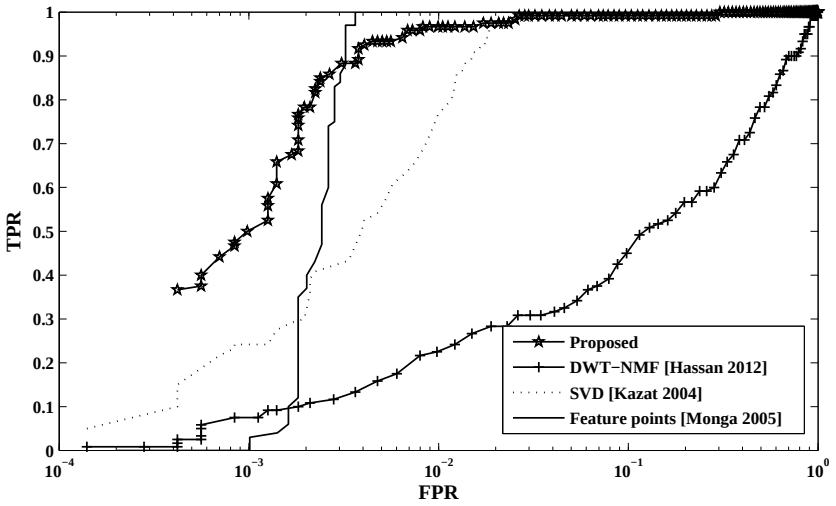


(a) AWGN

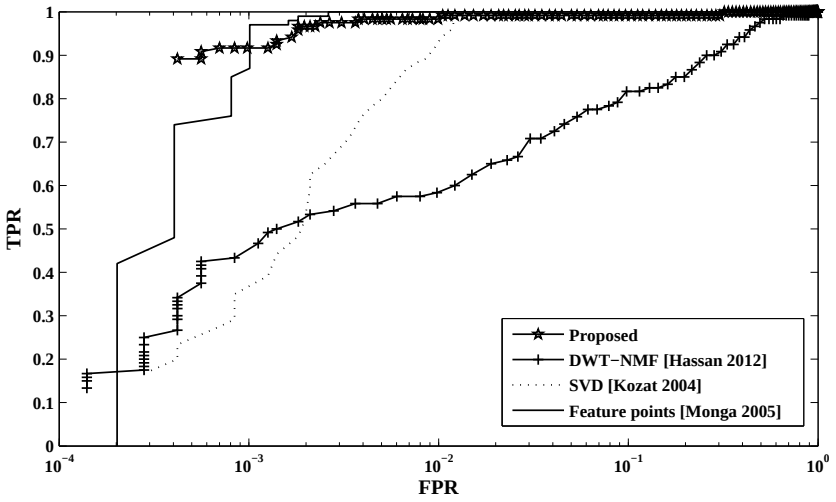


(b) Histogram equalisation

Figure 3.14: ROC curves for each type of manipulations when applying into different image hashing techniques, (a) AWGN (b) Histogram equalisation



(a) Rotation



(b) Translation

Figure 3.15: ROC curves for each type of manipulations when applying into different image hashing techniques, (a) Rotation (b) Translation

3.4.3 Unpredictability testing

In addition to robustness of image hashing, the security in terms of unpredictability bits that arises from the key-dependent randomisation is an important property of the image hashing technique. A high amount of randomness in the hash values makes it difficult

for the adversary to estimate or forge the hash values without knowing the secret keys. We have estimated correlations between different binary hashes via the binomial distribution (Daugman, 1993). We generated 120 hashes for the images in the database, and calculated the 7140 normalised Hamming distances between the hash pairs of different images. The distribution of the normalised Hamming distances is shown in Figure 3.16. It can be found that the distribution of the normalised Hamming distance corresponds to mean and standard deviation being $\mu = 0.492$ and $\sigma = 0.072$. Since the standard deviation of a binomial is given by $\sigma = \sqrt{p(1-q)/N}$ (where $p = 0.5$ and $q = 1-p$), this distribution of normalised Hamming distance would correspond to a binomial process where $N = 48$. A theoretical plot of the binomial process with $N = 48$ and $p = 0.5$ is also displayed by as a solid line in Figure 3.16. Therefor, the likelihood of two binary hashes from different images matching completely by chance is one in 2^{48}, or approximate 2×10^{-15}. This means that 48 out of 320 hash bits (15%) are independent and unpredictable.

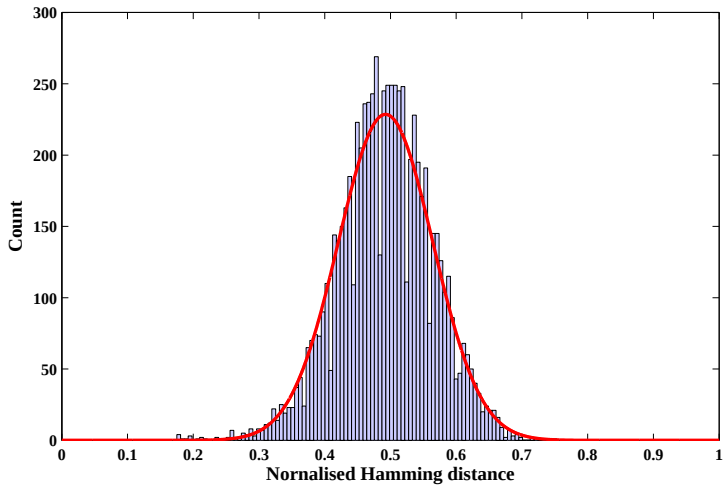


Figure 3.16: Distribution of the normalised Hamming distance between hashing pairs of different images

3.5 Summary

In this chapter, we introduced a new robust and secure image hashing technique using pseudo-randomly selected sub-images in discrete wavelet transform (DWT) and non-negative matrix factorisation (NMF). Based on our experimental results, it has been noted that the proposed technique is robust to a large class of content-preserving operations (CPOs). We compared different techniques e.g.feature points, SVD-based, and the DWT-NMF-based. Because, the proposed image hashing technique uses wavelet transform, it still suffers from some geometric distortions, such as rotation and translation. The non-negative matrix factorisation (NMF) dimension reduction technique is based on the approximate NMF, which factorises the image matrix into two lower rank matrices. Consideration is required in choosing a low rank r e.g. r_1, r_2 in the NMF because this effects the robustness and discriminability of the system. A bigger rank will make better image hashing performances. This is because of the increased the rank of decomposition means more elements in the hash, which not only preserve the perceptual robustness, but also improve the discriminative capability. Generally, a short image hash will have good robustness, but low discrimination. As hash length increases, discriminative capability is strengthened while perceptual robustness slightly decreases. Base on the basic properties of image hashing algorithm, note that image hash is a compact representation, meaning that hash length is expected to be short enough. Therefore, a trade-off is needed in choosing the hash length, i.e., the values of L and r. In the next chapter, we plan to explore the invariant transform referred Fourier-Mellin Transform (FMT) to extract robust features under content-preserving operations (CPOs) especially geometric attacks.

Chapter 4

Perceptual image hashing in Fourier-Mellin Transform (FMT) domain

The Fourier-Mellin transform (FMT) has been successfully used in numerous image recognition and registration applications, because it is invariant to rotation, translation and scaling (Lin et al., 2001; Alghoniemy M.and Tewfik, 2004; Guo et al., 2005; Swaminathan et al., 2006). Inspired by the potential of the FMT for image hashing, a new robust and secure image hashing technique based on overlapping blocks in FMT domain is introduced (Prungsinchai et al., 2013). The basic idea of the proposed image hashing technique consists of exploiting the properties of Fourier-Mellin transform into overlapping blocks to extract robust features. To secure the image hashing system, two secret keys were used in the pre-processing and hash computation steps. The rest of the chapter is structured as follows: The proposed framework for perceptual image hashing is described in section 4.1. Section 4.2 discusses the identification and evaluation measure of the image hashing algorithm. Experimental results for robustness testing, robustness versus discriminability testing and unpredictability testing are presented in section 4.3. Finally, the key ideas introduced in this chapter are summarised in section 4.4.

4.1 Proposed of perceptual image hashing in FMT domain

4.1.1 Fourier-Mellin Transform basic

As mentioned in section 2.2, let f denote a gray-scale level image defined over a compact set of $\mathbb{R}^2$. Here, the FMT can be divided into three steps as follows:

- ***The Fourier transform (FT)***: It converts the translation of the original image in the spatial domain.

$$I(x, y) \rightarrow |F\{I[m, n]\}| \tag{4.1}$$

  where x and y are cartesian coordinates.

- ***The cartesian to Log-Polar coordinates***: It converts to Log-polar coordinates, and then maps the scaling and rotation into the horizontal and vertical translations.

$$F[k, l] \rightarrow G(log\rho, \theta) \tag{4.2}$$

  where ρ and θ are Log-Polar coordinates.

- ***The Mellin transform***: Applying second Fourier transform (called Mellin transform) in Log-Polar coordinates and converts into the offsets of angles in the spectrum domain to converts the Log-polar coordinates image, and then to return the magnitude feature image. The output is invariant to rotation, scaling and translation.

$$F[u, v] \rightarrow F\{G(log\rho, \theta)\} \tag{4.3}$$

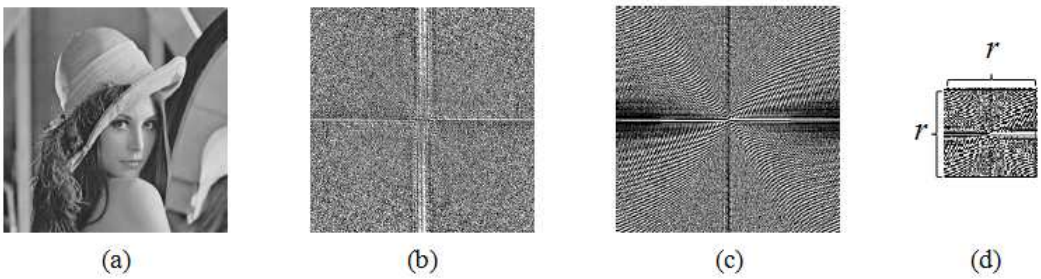


Figure 4.1: Example of Fourier-Mellin Transform, (a) Input image (b) Fourier spectrum (c) Log-polar of spectrum (d) low-frequency area with size r-by-r

The proposed technique follows a three steps framework to generate the hash. As illustrated in Figure 4.2.

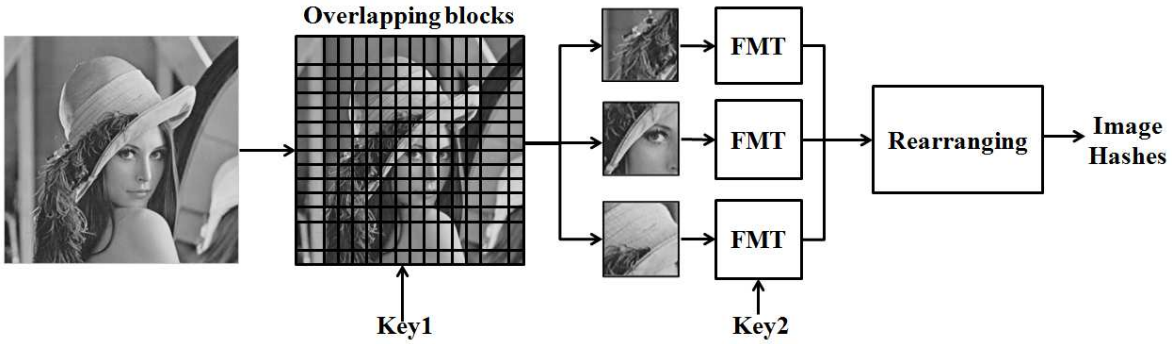
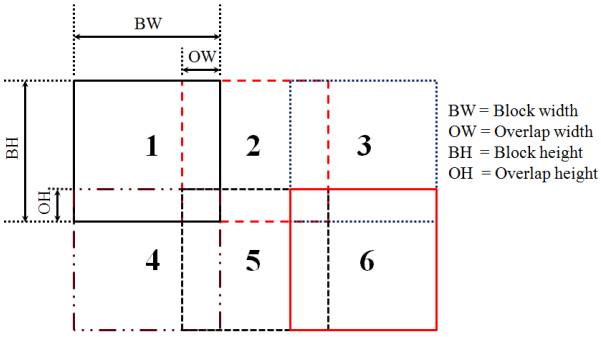


Figure 4.2: The proposed FMT-based image hashing scheme

These three steps include:

- ***The pre-processing step***: Given an input image I with size M-by-M. A low pass filtering by Gaussian kernel with a zero mean and $\sigma = 0.5$ variance is applied, and then a histogram equalisation is performed on the input image. This step aims to reduce the effect of the common signal processing operations. The image is divided into overlapping blocks size of m-by-m with the horizontal and vertical overlapping of n-by-n, $I_{block1}, I_{block2}, \ldots, I_{blockN}$ as shown in Figure 4.3 (a) and (b). The indices of the block sequence $I_{block1}, I_{block2}, \ldots, I_{blockN}$ are randomly scrambled using a secret key K_1 to obtain a block sequence with a new scanning order $I'_{block1}, I'_{block2}, \ldots, I'_{blockN}$.

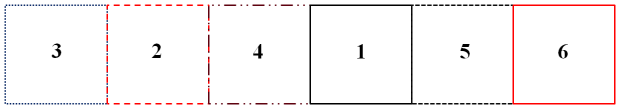
An example of randomly ordered overlapping block with different secret keys is shown below in Figure 4.4.



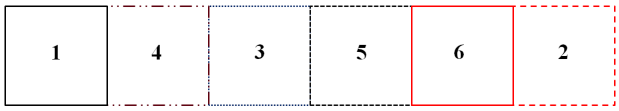
(a) Overlapping blocks

(b) "Lena" image with blocks size 256-by-256 overlapping 128-by-128

Figure 4.3: Example of the overlapping technique



(a) Secret key K_1



(b) Secret key K_2

Figure 4.4: Example of a random ordered overlapping blocks with different secret keys

- ***The feature extraction step***: FMT is applied to I'_{blockN} to extract features. In this step, the central part of the low frequency area with size r-by-r is selected, denoted by F'_{blockN}, as illustrated in Figure 4.1 (d).
- ***The hash computation step***: Each low frequency matrix from F'_{blockN} is converted

to vector $V_1, V_2, \ldots, V_N$ (excluding the symmetric part of the spectrum). Then, two low frequency features from each V_N are randomly selected depending on the another secret key, K_2. Next, the final hash vector V_F is formed by concatenation of the low-frequency features.

4.2 Identification and similarity measure

The evaluation of the proposed image hashing algorithm is conducted in two aspects: identification accuracy and Receiver Operating Characteristics (ROC) analysis.

4.2.1 Identification process

The Euclidean distance is applied as a performance metric to measure the similarity and discriminating capability between two hash vectors. The lower the Euclidean distance is, the closer the two hash values. Let H_1 is the hash of the original image and H_2 is the hash of a similar version or different from the original image and it is defined as:

$$ED((H_1),(H_2)) = \sqrt{\sum_{i=1}^{n}(h_1(i) - h_2(i))^2} \tag{4.4}$$

Normalised Euclidean distance (NED):

The Euclidean distance can be normalised with respect to the maximum distance of the Euclidean distance strings, then normalised in the [0,1] range. The two images are perceptually similar and the distance is close to 0, whereas the distance is expected to be close to 0.5 for two distinct images. The normalised Euclidean distance is defined as:

$$NED = \frac{ED}{\Lambda} \tag{4.5}$$

where Λ is the maximum Euclidean distance among all tested distances.

4.2.2 Receiver Operating Characteristics analysis

To obtain the ROC curve, the $TPR(\tau)$ and $FPR(\tau)$ are defined as:

$$TPR(\tau) = Probability(D(H(I,K), H(\hat{I},K)) < \tau) \tag{4.6}$$

$$FPR(\tau) = Probability(D(H(I,K), H(J,K)) < \tau) \tag{4.7}$$

where τ is the identification threshold. The image $\hat{I}$ is a modified version of I and J is a distinct image of the original image I. Based on all distances between the original images and their attacked versions. ROC curves were generated by varying the threshold τ from the minimum to the maximum value of all the distances. Given a certain threshold τ, a better image hashing should plot a higher TPR(τ) with a lower FPR(τ), simultaneously.

4.2.3 Database and content-preserving operations

The dataset with 120 original gray scale natural images were constructed. For each original image with size 512×512, the similar versions were generated by manipulating the original image according to a six classes of content-preserving operations as listed in Table 4.1. The motivation is to construct such a database was to simulate some possible quality change or a different format of digital images due to noise, lossy compression and small rotation manipulation. The normalised Euclidean distance between the hashes of the original image and their attacked versions were measured. The normalised Euclidean distance between the hashes of dissimilar images were also considered, which indicated the discriminative capability of the hashing algorithm.

Table 4.1: Content-preserving operations with various parameters

Manipulation type	Parameters
Image processing operations	
JPEG compression	quality factor $QF = 10\sim90$
Additive White Gaussian Noise (AWGN)	standard deviation $\sigma = 20\sim35$
Median filtering	window size 3∼9
Histogram equalization	/
Geometric distortions	
Rotation	degree 5°∼9°
Translation	window size 5∼20

4.3 Experimental results with the proposed image hashing technique in FMT domain

4.3.1 Robustness testing

This is to examine the robustness of the algorithm designed in section 4.1. The scheme is mainly designed to successfully handle the rotation attacks and the trade-off between the robustness and security. Therefore, the performance of FMT image hashing is tested with respect to the difference of block size and overlapping size under content-preserving operations such as JPEG lossy compression, median filtering, rotation and so on. For the FMT approach, the parameters are set as shown in Table 4.2. The overlapping blocks can be an effective approach to enhance security. The use of overlapping blocks can be justified by their robustness under geometric changes (Khelifi and Jiang, 2010). For the sake of illustration, Table 4.3 depicts results on the robustness against a number of geometric attacks measured by the normalised Euclidian distance for "Lena" image with regards to the use of overlapping and non-overlapping blocks. It was clearly observed that the features extracted from non-overlapping and overlapping blocks did not have any significant effect.

In the rest of this chapter, the block size 64 overlapping size 16×16 with horizontal and

vertical like squares of the same size and low frequency area $r = 10$ will be adopted. The performance of the image hashing algorithm for three example images namely "Lena", "Baboon" and "Peppers" are shown in Figures 4.5 to 4.7. Figure 4.5 (a) shows the performance of the hashing algorithm under JPEG lossy compression. As can be seen, the normalised Euclidian distance between two hashes extracted from the original image and compressed image with a quality factor up to 10 is 0.34. The results obtained under JPEG lossy compression show a good performance when the images are compressed with a quality factor higher than 30%. The normalised Euclidian distance under median filtering and AWGN attacks are lower than 0.45 and 0.35 as illustrated in Figures 4.5 (b) and 4.6 (a), respectively. This is the effect of incorporating the low pass filter and histogram equalisation in the pre-processing step.

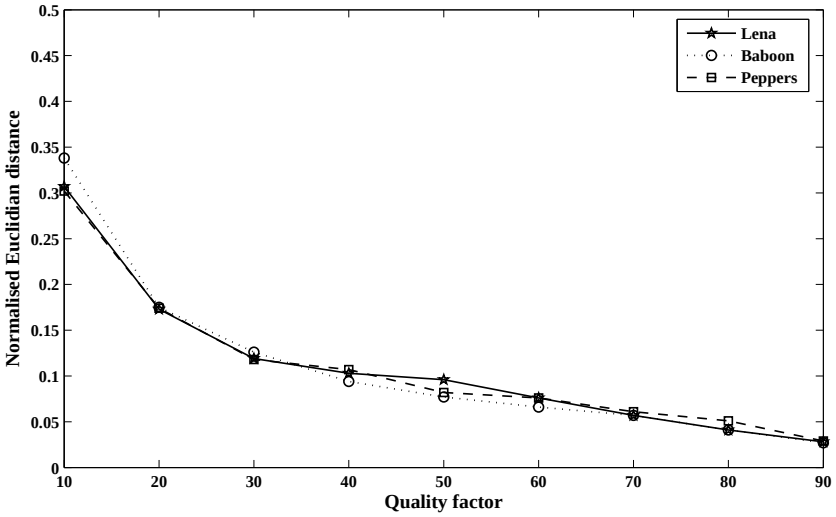
In the case of geometric distortions, it can be observed that the performance of the proposed technique is reasonably good for both attacks (see Figures 4.6 (b) and 4.7). In Figures 4.6 (b) and 4.7 the normalised Euclidian distance between the hashes of the distorted images and original images under rotation and translation attacks are close to 0.3 and lower than 0.35, respectively. It can be noted that the rotation and translation invariance properties of Fourier-Mellin transform makes the system robust under these attacks, unlike other transforms such as the wavelet transform seen in chapter 3.

Table 4.2: Parameters setting in the FMT-based image hashing algorithm

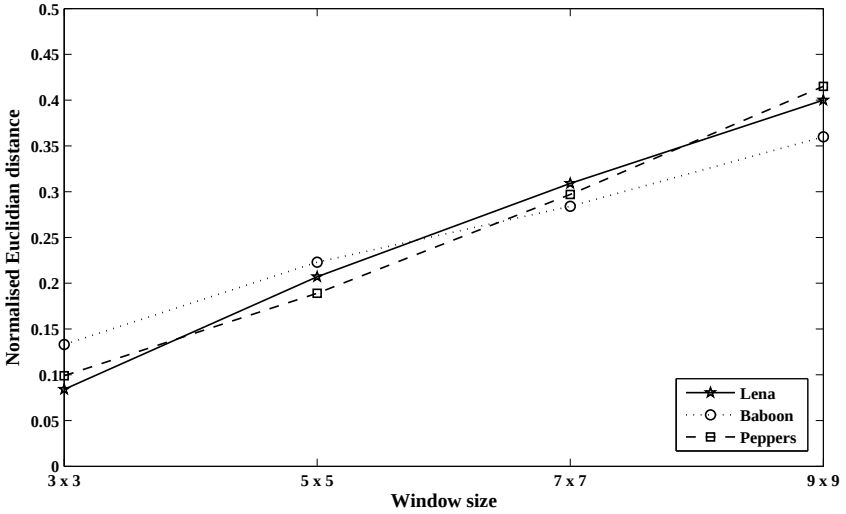
Parameter	**Value**
Size of the block	$m = 64$ and 32
Size overlapping	$n = 16$
Length of the hash vector	128, 200, 512 and 1922
Low frequency area	$r = 10$

Table 4.3: Normalised Euclidian between the feature vectors extracted from the original "Lena" image and its attacked versions Ov.:Overlapping blocks by sixteen pixels Non-Ov.: Non-overlapping blocks

Attacks	**Block size 32**		**Block size 64**	
	Non-Ov.	**Ov.**	**Non-Ov.**	**Ov.**
Hash length	512 real value	1922 real value	128 real value	200 real value
Rotation 3°	0.106	0.110	0.128	0.101
Rotation 5°	0.164	0.165	0.152	0.162
Rotation 7°	0.213	0.212	0.202	0.214
Rotation 9°	0.252	0.249	0.248	0.253
Rotation 10°	0.263	0.262	0.269	0.269
Translation 3×3	0.090	0.091	0.093	0.095
Translation 5×5	0.153	0.151	0.151	0.154
Translation 7×7	0.211	0.209	0.206	0.208
Translation 9×9	0.260	0.261	0.261	0.260
Translation 10×10	0.284	0.286	0.288	0.282

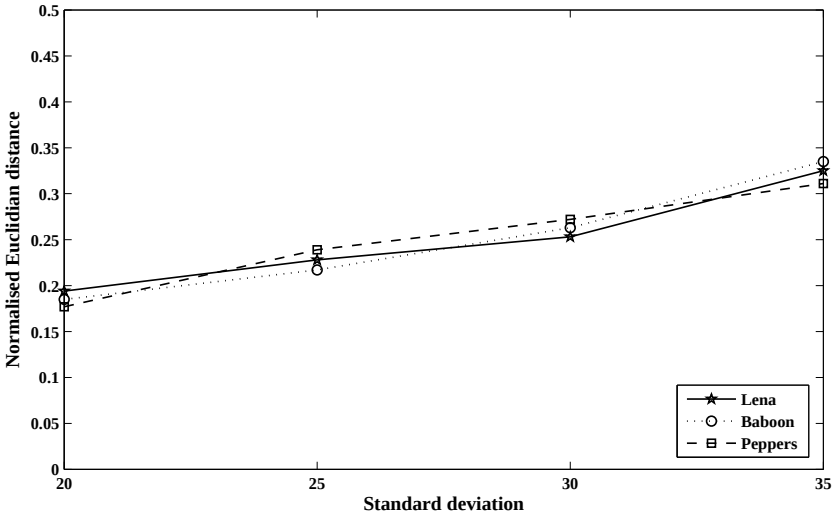


(a) JPEG lossy compression

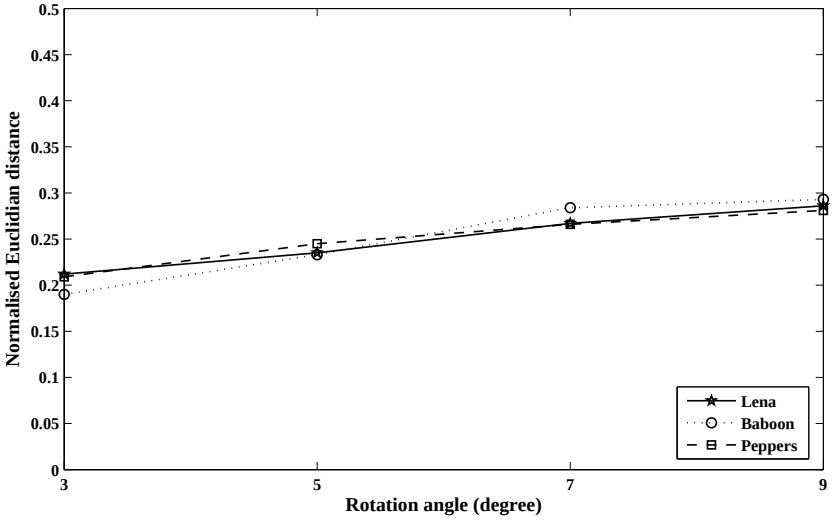


(b) Median filtering

Figure 4.5: Performance robustness of the proposed technique under (a) JPEG lossy compression (b) Median filtering



(a) AWGN



(b) Rotation

Figure 4.6: Performance robustness of the proposed technique under (a) Additive white Gaussian noise (b) Rotation

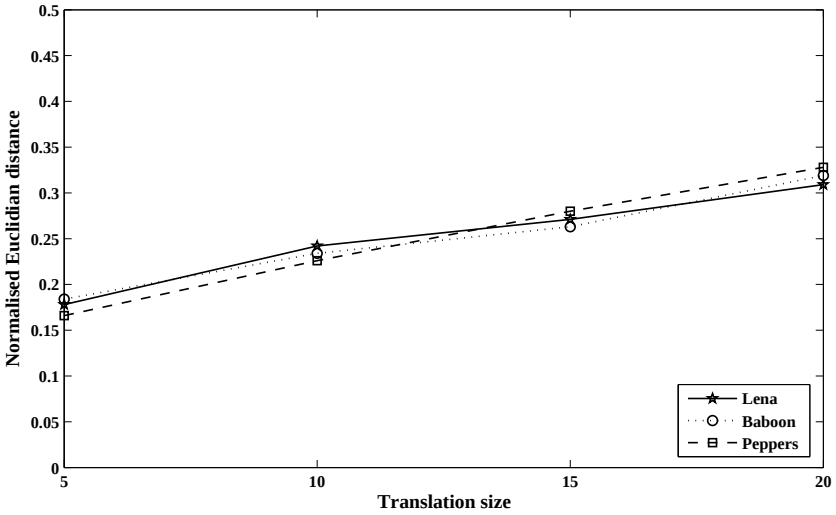


Figure 4.7: Performance robustness of the proposed technique under Translation

4.3.2 Robustness versus discriminability

Recall from section 1.2 that the properties of perceptual image hashing 1 and 2 determine the performance in terms of robustness and differentiation. To evaluate both of these properties, a statistical comparison between the proposed image hashing technique and different image hashing algorithms has been con through ROC curves. Given the distance between visually different images $\hat{d}_i$ (i = 1, 2,..., q), and the distance between hashes extracted from the original and the attacked images, d_i (i = 1, 2,..., s). Based on 120 images, $q = 7140$ possible distances can be obtained, and by considering ξ different attacks as listed in Table 4.1, are divided in two classes: *image processing operations* and *geomantic attacks*, $s = 120 \times \xi$. Image processing operations include JPEG lossy compression with $QF = 10$, AWGN with standard deviation $\sigma = 20$, median filtering with window size 3×3, histogram equalisation, while geometric attacks are rotation with 5° degree and translation with window size 5×5. To show the advantages of the proposed hashing technique and make the comparison as fair as possible, a number of hashing algorithms were also applied on the same images and with the same attacks; namely wavelet-based hashing (Venkatesan et al., 2000), feature points-based image hashing (Monga et al., 2005) and sub-images-

DWT-based image hashing (Prungsinchai et al., 2012). The algorithms parameters used in this implementation are described in Table 4.4. Table 4.5 lists the hash length obtained for each image hashing technique. It is worth mentioning that the distance used to compute TPR and FPR depends on the nature of the hash. Indeed, the wavelet-based image hashing and sub-images-DWT-based image hashing techniques extracts binary hashes, and hance the normalised Hamming distance (NHD) is used. The Euclidean distance is used for real valued hashes extracted via proposed image hashing technique. Finally, the Hausdorff distance is used for hashes corresponding to the coordinates of feature points extracted via the feature points image hashing technique. The parameter used in implementation and the hash length obtained from each image hashing techniques are listed in Table 4.4 and 4.5, respectively.

Table 4.4: Parameters used in the implementation Ov.:Overlapping blocks

Hashing technique	**Parameters**
Proposed	block size 64 Ov. 16 pixels and $r = 10$
Wavelet (Venkatesan et al., 2000)	wavelet 3 levels, N = 150 rectangles
Sub-images and DWT (Prungsinchai et al., 2012)	$P = 20$, $m = 128$, and $r = 1$
Feature points (Monga et al., 2005)	64 feature points

Table 4.5: Hash length for each assessed technique

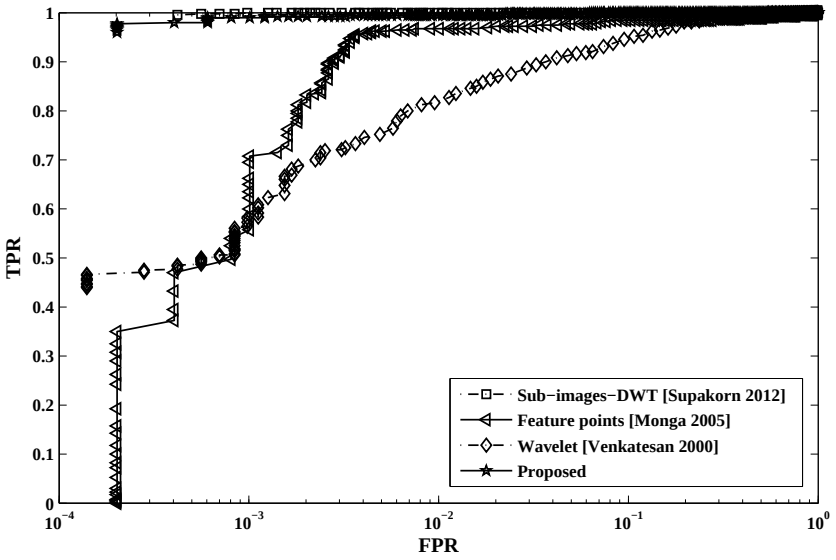
Hashing technique	**Hash length**
Proposed	200 (real values)
Wavelet (Venkatesan et al., 2000)	600 (binary)
Sub-images and DWT (Prungsinchai et al., 2012)	320 (binary)
Feature points (Monga et al., 2005)	64×2 = 128 (coordinates)

The overall ROC curves for all types of tested manipulations when applying different hashing scheme is shown in Figure 4.8. Figure 4.8 (a) if the feature points image hashing techniques is taken as a reference point, we can be seen that the both the sub-images-DWT image hashing technique and the proposed image hashing technique perform equally

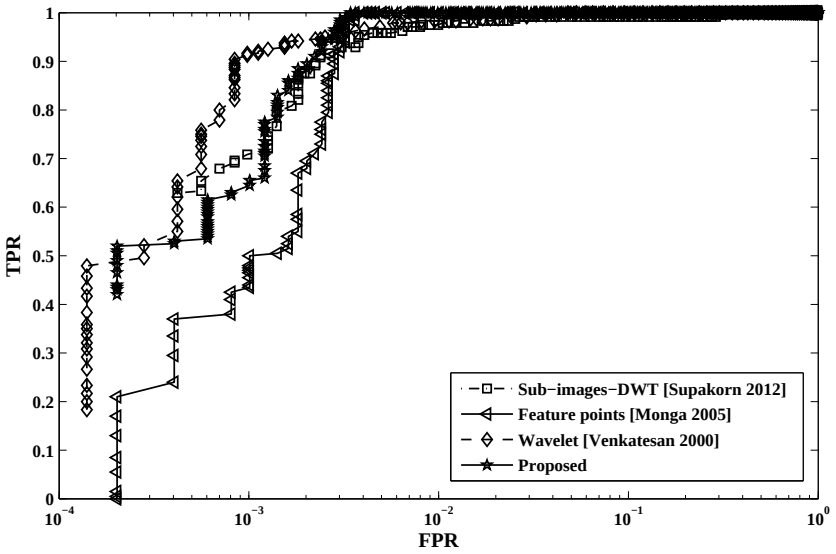
well under image processing operations. Once again, from the ROC curves in Figure 4.8 (b), it is clear that the proposed image hashing technique significantly outperforms other state-of-the art techniques in a reasonably good range for TPR > 0.950, FPR $> 10^{-2.548}$. This can justified by the fact that the coefficients extracted from the FMT-transform resist geometric attacks as mention in chapter 2.

To show the robustness to each type of attacks, an ROC curve is also generated for each attack and hash algorithm are shown in Figures 4.9 to 4.11. Figures 4.9 (a) and (b) illustrates the image hashing performance under JPEG lossy compression and median filtering. This is because the generated signatures from the coefficients of the LL subband are invariant against such attacks. Figures 4.10 (a) and (b) show the performance under histogram equalisation and AWGN. It can be seen that the proposed image hashing technique give superior performance due to the pre-processing step resisting under these attacks, Obviously, wavelet-based image hashing provides less robustness under AWGN attack, because the feature coefficients are sensitive to such attack. While feature-point technique performed better under AWGN attack.

In Figure 4.11 (a) the proposed image hashing technique provides the best performance under rotation attack in the range of TPR > 0.895 and FPR $> 10^{-2.475}$. Moreover, the proposed image hashing provides the best performance under translation attack as shown in Figure 4.11 (b).

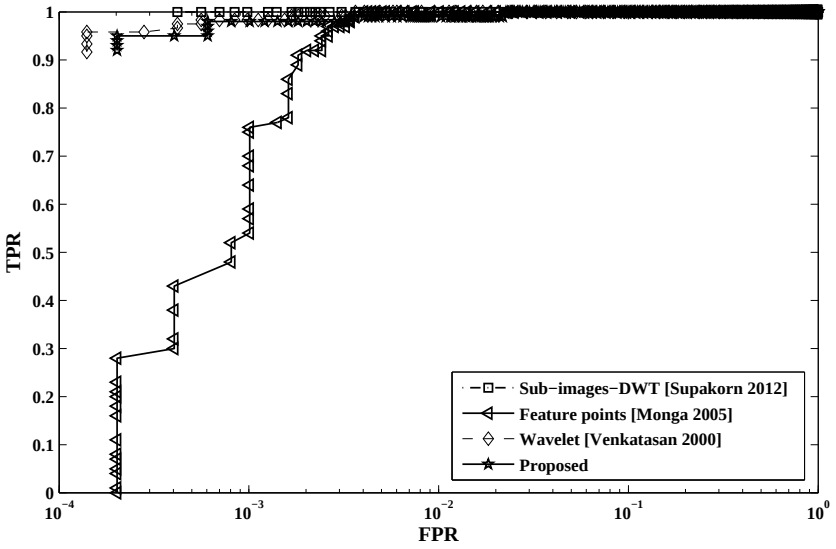


(a) Image processing operations

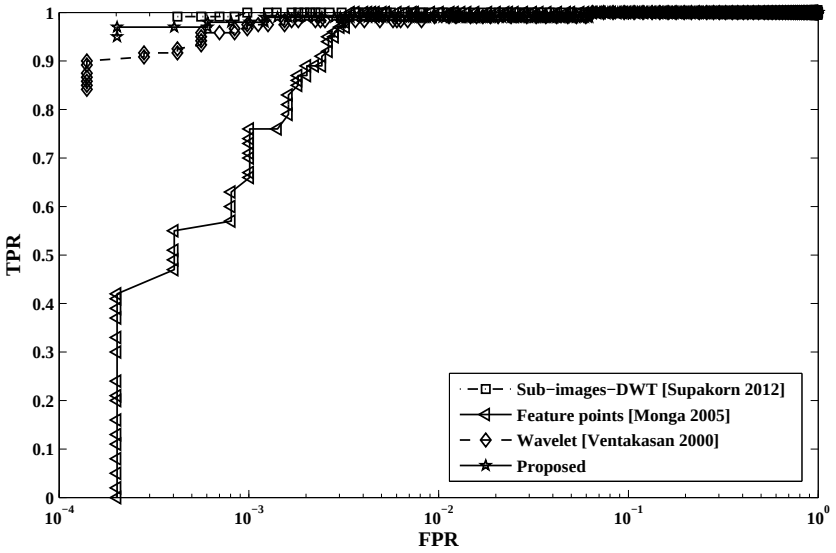


(b) Geometric attacks

Figure 4.8: The overall ROC curves for all types of test manipulations when applying different hashing schemes, (a) Image processing operations (b) Geometric attacks

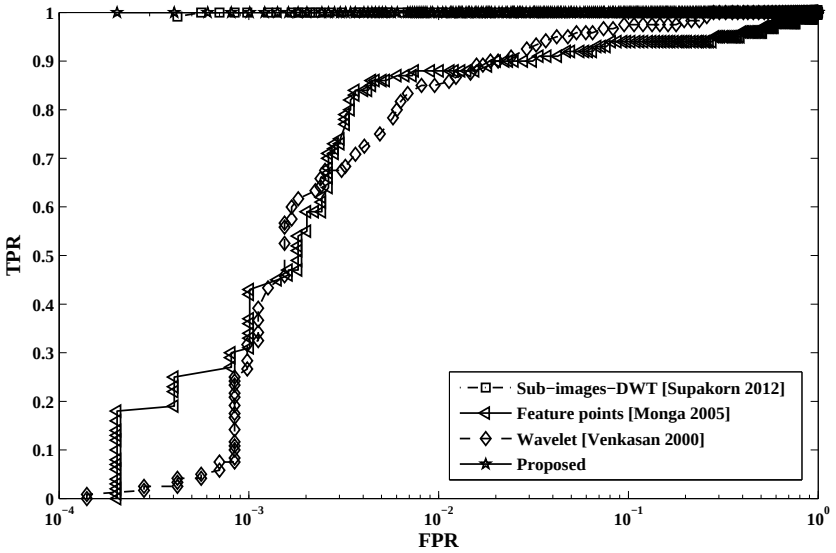


(a) JPEG lossy compression

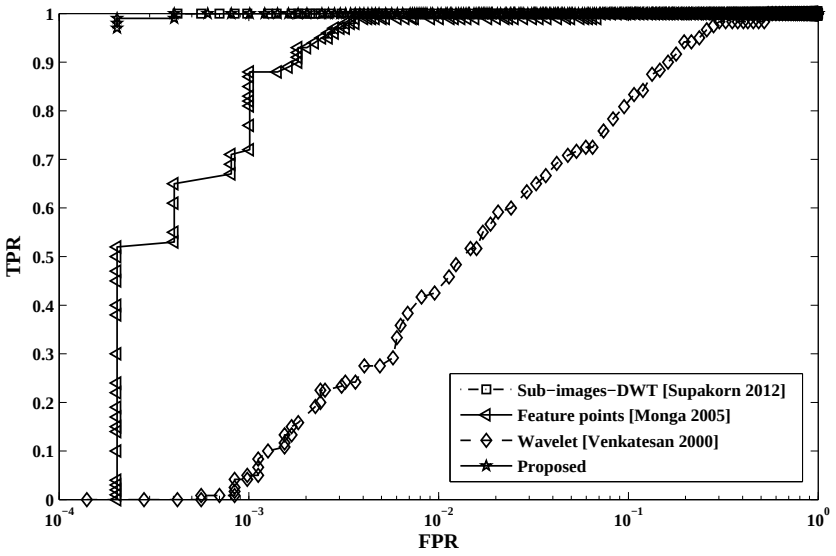


(b) Median filtering

Figure 4.9: ROC curves for each type of manipulations when applying into different image hashing schemes, (a) JPEG lossy compression (b) Median filtering

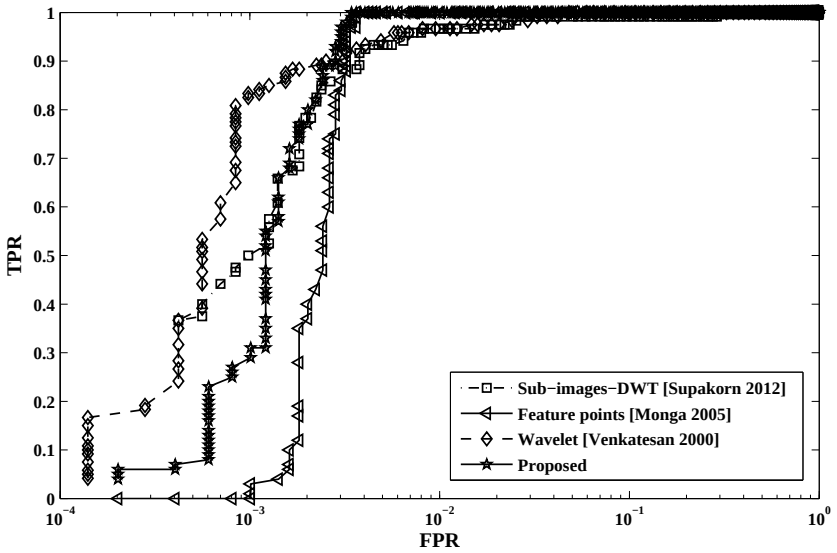


(a) Histogram equalisation

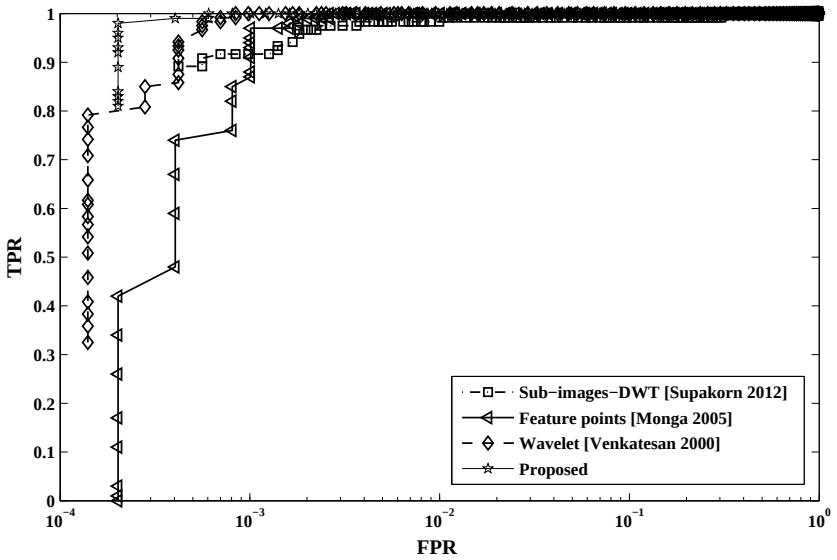


(b) AWGN

Figure 4.10: ROC curves for each type of manipulations when applying into different image hashing schemes, (a) Histogram equalisation (b) AWGN



(a) Rotation



(b) Translation

Figure 4.11: ROC curves for each type of manipulations when applying into different image hashing schemes, (a) Rotation (b) Translation

4.3.3 Unpredictability testing

It is extremely important to ensure that the output produced by a hashing system cannot be estimated or forged without knowing the correct secret key. Here, we have estimated correlations between different hashes via the binomial distribution (Daugman, 1993). Ideally, if each hash is fully independent of every other hash, then the distribution of the normalised Euclidean distances between such independent hashes is binomial with $p = 0.5$ and $N = 200$. The actual distribution of 7140 observed normalised Euclidean distances between hashes extracted from different image is shown in Figure 4.12. The empirical distribution has a standard deviation $\sigma = 0.095$ and with a mean $\mu = 0.528$. Note that most of the hash distances between different images are closed to 0.5, which is close to the ideal situation. Since the standard deviation of a binomial distribution is given by $\sigma = \sqrt{p(1-q)/N}$, the distribution of hash distances corresponds to a binomial process where $N = 28$. A theoretical distribution is plotted of the binomial distribution with $N = 28$ and $p = 0.5$, and also displayed by a solid line in Figure 4.12. As can be seen, the theoretical distribution approximately fits the actual data. Therefore, the likelihood of two hash from different images matching completely by chance is one in 2^{28}. That is, 28 out of 200 hash (14%) are independent and unpredictable. This shows a limitation of the proposed technique which is mainly due to the nature of FMT. Indeed, the FMT is not an efficient decorrelating transform unlike discrete cosine transform (DCT)and the Karhunen-Loève transform (KLT) and hence the coefficients used to extract the hash bits are correlated with each other to some extent.

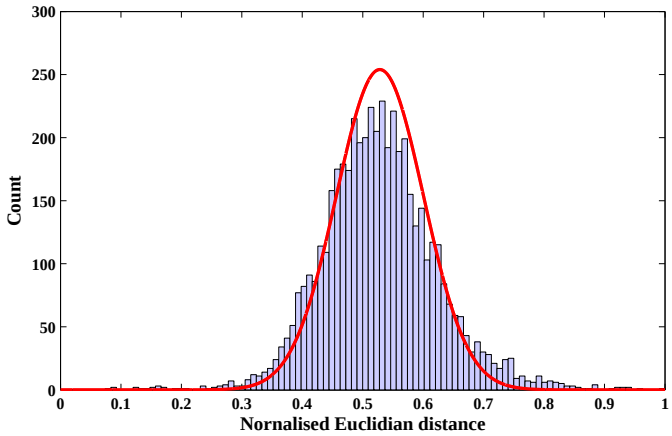


Figure 4.12: Distribution of normalised Euclidian distance between hashing pairs of different images

4.4 Summary

In this chapter, a robust and secure perceptual image hashing technique based on the fourier-Mellin transform (FMT) is proposed. Based on our experimental results, it has been shown that the proposed FMT-based image hashing technique is robust to a large class of image processing operations. The invariance property to rotation, scaling and translation of FMT makes it more suitable for image hashing. Robust hashing is a promising solution to the content identification problem. The performance of the image hashing algorithm has been assessed via the ROC curves. The experiments showed that the proposed algorithm significantly outperforms state-of-the art algorithms by achieving a higher TPR, while it maintaining lower FPR. The robustness of the system was observed under geometric attacks except for large rotation and translation manipulations, and that the distance between the hash values of perceptually similar images was clearly separated from the distance between different images.

Chapter 5

Perceptual image hashing in Discrete Cosine Transform (DCT) domain

At present, the discrete cosine transform (DCT) is the current standard method for compression in JPEG and MPEG (Pennebaker and Mitchell, 1992; International Standard, ISO/IEC/JTC1/SC29 WG11, 1998). In 2002, a new wavelet-based compression standard JPEG2000 was introduced, however, JPEG2000 has not been able to supplant JPEG, which remains a main compression standard for digital cameras (Ponomarenko et al., 2007). Several DCT-based image hashing algorithms have been proposed (Lin and Chang, 2001; Sun et al., 2002; Kailasanathan et al., 2003; Tang et al., 2005), which have the advantage of strong features for the transform to provide robustness. These robust features should be robust/strong enough to minor pixel modifications that arise from image processing manipulations such as JPEG lossy compression, median filtering and so on. For example, Lin and Chang (2001) proposed a robust image authentication method distinguishing JPEG lossy compression from malicious manipulation. This was based on invariance in the relationship between DCT coefficients at the same position in separate blocks of image. Their method can distinguish malicious manipulations from JPEG lossy compression, irrespective of how high the compression ratio is. Kailasanathan et al. (2003) presented a DCT based hashing function which resists an acceptable level of compression and image processing operations such as Gaussian noise addition and median filtering by considering

a subset of DC coefficients together with number of AC coefficients. A robust image hashing technique in the DCT domain was proposed by Tang et al. (2005). This scheme used the DC and AC low frequency sub-bands, of DCT blocks to increase the signature's discriminability. The results from this scheme indicated that the similarity between original images and their corresponding JPEG lossy compression images were fairly high although the scheme was not robust against geometric manipulation.

Recently, the DCT sign coefficients have been used in image registration, content-based copy detection and image hashing (Kondo, 2001; Arnia et al., 2006; Yu and Sun, 2006; Arnia, Fujiyoshi and Kiya, 2007; Arnia, Iizuka, Fujiyoshi and Kiya, 2007*a*,*b*; Arnia et al., 2009). For instance, Yu and Sun (2006) presented image robust hashing based on DCT sign coefficients. In this algorithm, the DC sub-band of the input image was obtained using a HAAR wavelet. Rectangular blocks with size of 64×64 are sequentially selected for DCT calculating. Sign extraction was performed in the DCT coefficients matrix of each rectangular, then the first 32^{th} AC coefficients in zigzag order were selected. This algorithm robustness under a large class of content-preserving operations (CPOs), especially under the cropping attack which consists of removing rows and columns from the image.

Inspired by the potential of DCT for image hashing, two schemes of image hashing were introduced. The basic idea of both image hashing methods are using the low frequency sub-band of DCT coefficients to achieve of the hash function for content identification issues. Section 5.1 presents the proposed image hashing technique in DCT domain. The experimental results and analysis including robustness testing, robustness versus discriminability testing and unpredictability testing are demonstrated in section 5.3. Finally, section 5.4 concludes the key ideas introduced in the chapter.

5.1 Proposed of perceptual image hashing in DCT domain

In this chapter, we propose two image hashing algorithms: algorithm A-DCT overlapping block-based hashing and algorithm B-DCT sign-based hashing. The algorithm A is firstly

presented as it uses the low frequency DCT coefficients in blocks and a randomisation process to increase the security of the hashing system. Secondly, algorithm B, uses the sign of DCT coefficients, which carry information on edges and texture. Sign coefficients can be observed as the image hash for the content identification purpose.

5.1.1 Algorithm A-DCT overlapping block-based image hashing

As mentioned in section 2.3, DCT is widely used for data compression and many other operations. Owing to its ability to compact the energy of the image in a few coefficients, DCT clusters high valued coefficients in the upper left corner (low frequency range) and low value coefficients in the bottom right of the corner (high frequency range). The image could be represented by a few DCT coefficients without losing much of the image quality. The algorithm A-DCT overlapping blocks-based image hashing scheme, is illustrated in Figure 5.1. This method adopted the block images strategy and low frequency DCT coefficients of every image block as robust features. The process is detailed as follows:

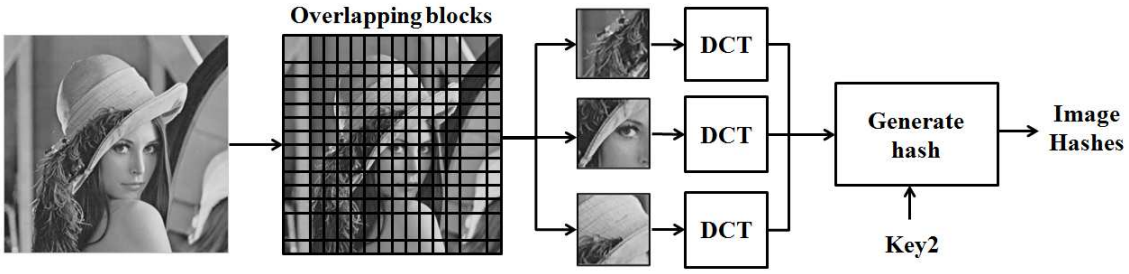


Figure 5.1: The proposed DCT overlapping block-based image hashing scheme

- ***The pre-processing step***: Low pass filtering of the input image I, this is to reduce noise. The low pass filtered image is divided the image into overlapping blocks with size m-by-m with the horizontal and vertical overlapping of n-by-n, denoted as $I_{L_{block1}}, I_{L_{block2}}, \ldots, I_{L_{blockN}}$, N is the number of blocks.

- ***The feature extraction step***: DCT is applied to each block $I_{L_{blockN}}$ to extract the features. There are three frequency coefficient sets: low frequency sub-band, mid

frequency sub-band and high frequency sub-band. Each block is read in a zigzag fashion to obtain the new DCT sequence as a vector (see Figure 5.2.(b)). Here, the image hashing algorithm is based on the fact that much of the signal energy lies in the low-frequency range around the top left corner, which contains the most important visual parts of the image (see Figure 5.2). The first 2 AC coefficients are pickup from each block. The DC coefficients contain the lowest frequency information of the input image block, while the AC coefficients contain the detail information. The output is formed by concatenating the selected AC coefficients in a new vector, V'.

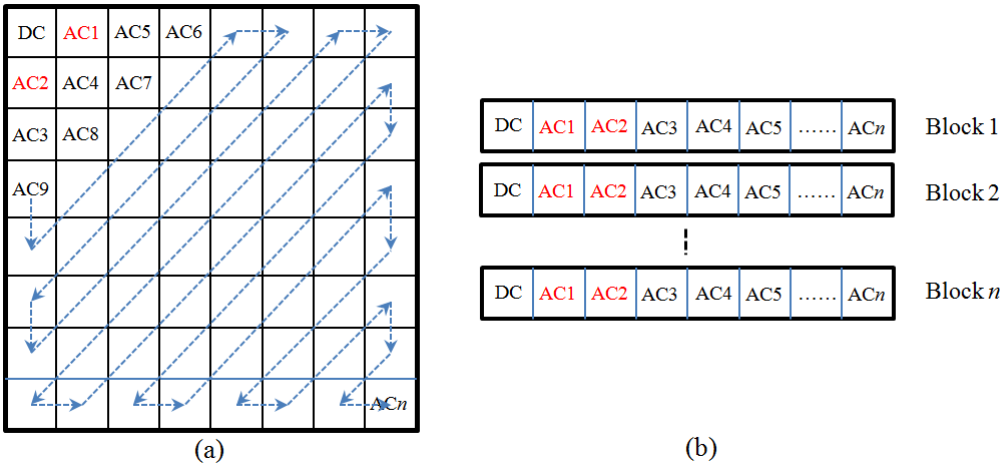


Figure 5.2: Example feature extraction step, (a) Original data matrix (b) Read zigzag order into a vector

- ***The hash computation step***: This is the process of extracting the binary hash via binarisation. A secret key is first used to pseudo-randomly arrange the AC coefficients in new vector R_i, $i = 0,1,\ldots,L$. Then the median value M_d is obtained in this sequence. Note that the use of randomness is important for security proposes.

$$M_d = median(R_i)(i = 1, 2, \ldots, N) \tag{5.1}$$

Then the hash is obtained in binary from as follows:

$$h(i) = \begin{cases} 0,\ R_i < M_d \\ 1,\ R_i \geq M_d \end{cases}$$

5.1.2 Algorithm B-DCT sign-based image hashing

As mentioned in section 2.3 Eq. 5.2 the DCT is given by:

$$C(k,\ell) = \alpha(k,\ell) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos(\frac{(2i+1)k\pi}{2N}) \cos(\frac{(2j+1)\ell\pi}{2N}) \tag{5.2}$$

where $f(i,j)$ is the spatial image and $\alpha(k,\ell)$ is the DCT coefficient. Gain control $\alpha(k,\ell)$ is given by:

$$\alpha(k,\ell) \begin{cases} \frac{1}{\sqrt{2}} = (k = 0) \\ 1 = (k \neq 0) \end{cases}$$

Only the sign values of the DCT coefficients are utilised. The sign of $C(k,\ell)$ are taken out and inspired by DFT coefficients sign $S(k,\ell)$, given by:

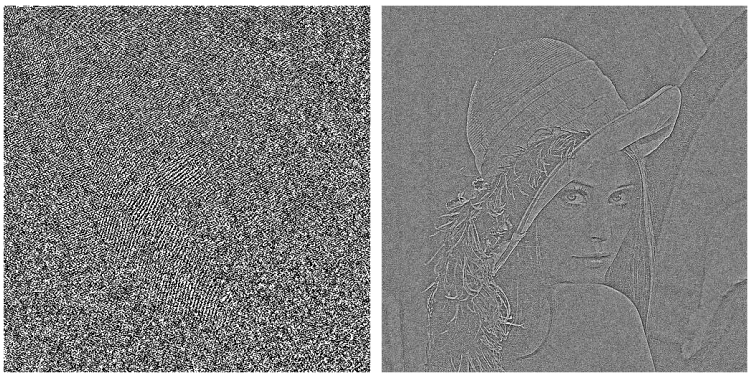
$$S(k,\ell) = sgn(C(k,\ell)) \tag{5.3}$$

Here sgn(x) is determined using the following equation

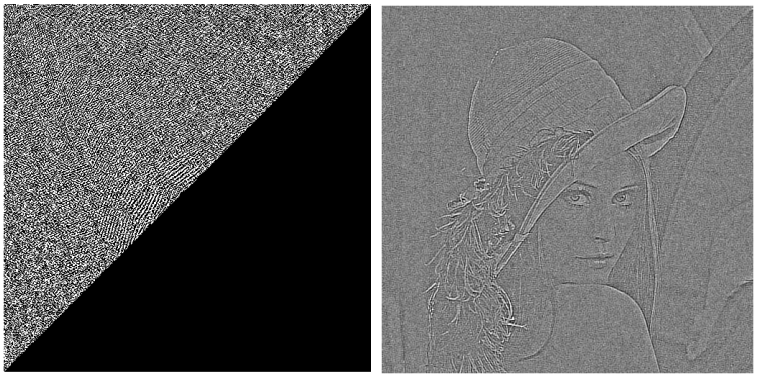
$$sgn(x) \begin{cases} -1 = (x < 0) \\ 1 = (x > 0) \end{cases}$$

The inverse DCT (IDCT) of the sign is called the DCT sign only image (DSOI) (Arnia et al., 2006), and is obtained by applying the 2D IDCT to $S(k,\ell)$. Although the DSOI is visually different from the original image (see Figure 5.3.(b)), it represents well the main structure, e.g. texture, edges, contours, etc. and hence can be distinguished from other images. Due to the compressive nature of the DCT, it is not necessary to use all DCT sign coefficients, in order to obtain a high-quality representation of the original image.

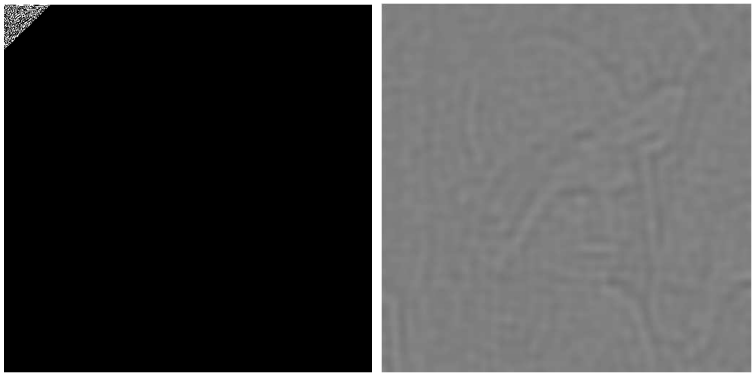
Generally, half the number of the DCT sign coefficients, taken in a zigzag order can be used to reconstruct the DSOI of the original image (see Figure 5.3.(c)). As can be seen in Figure 5.3.(e), by observing the DSOI reconstructed with only 2048 numbers of the DCT sign coefficients main part in relation to information that the original image carries can be observed and allows to recognise the original image. Figure 5.4 showed the example of zigzag and inverse zigzag order technique.



(a) Total number of sign bits (b) Reconstructed image of (a)

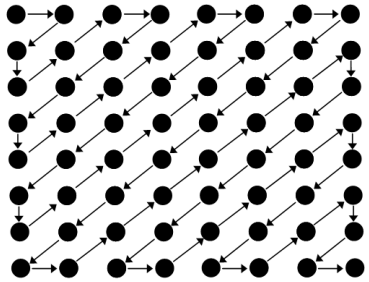


(c) Half number of sign bits (d) Reconstructed image of (c)

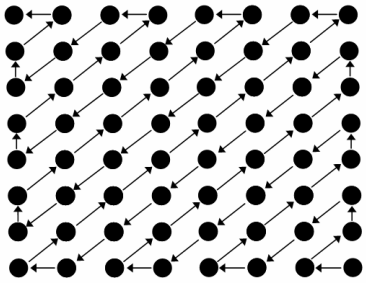


(e) Only 2048 of sign bits (f) Reconstructed image of (e)

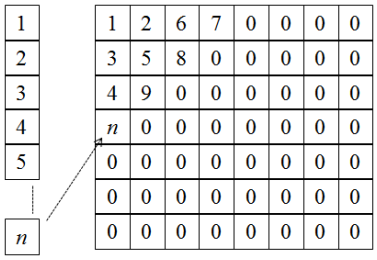
Figure 5.3: Reconstruction of the DCT sign only image (DSOI) for "Lena"



(a) Zigzag order



(b) Inverts zigzag order



(c) Inverse zigzag with added zeroes

Figure 5.4: For example zigzag and inverse zigzag order

The algorithm B-DCT sign-based image hashing scheme is illustrated in Figure 5.5. The main idea is to use the energy compaction property of the DCT sign values that carry a significant part of information representing the image in terms of texture, contours and the edges of areas as perceptual features. The procedure consists of the following steps:

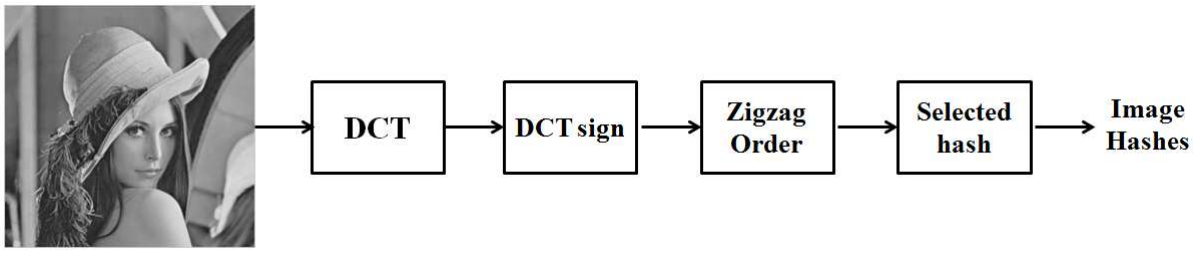


Figure 5.5: The proposed DCT sign-based image hashing scheme

- ***The feature extraction step***: Given an input image I. DCT is applied onto I to extract the feature coefficients. Then sign extraction in Eq. 5.3 performed on DCT coefficients matrix.

- ***The hash computation step***: After the feature extraction step, all sign coefficients matrix were read in zigzag-scanned to form vector, and then the sign coefficients from the first to the n^{th} AC coefficients are selected to be the image hashes to represent to the original image. A negative value of sign coefficient corresponds to a hash bit of zero (0). Likewise, a positive sign value corresponds to hash bit 1.

5.2 Identification and similarity measure

5.2.1 Identification process

The evaluation for the perceptual robustness of the proposed of image hashing algorithms are to conduct the assessment in two aspects: identification accuracy and ROC analysis.

Algorithm A-DCT overlapping block-based image hashing

The normalised Hamming distance (NHD) can be used to measure the similarity between two binary image hashes in algorithm A. Let $h_1(i)$ be the binary hash of the original image and $h_2(i)$ be the binary hash of a similar version or different from the original image. The normalised Hamming distance can be estimated between H_1 and H_2 and defined as:

$$NHD((H_1),(H_2)) = \frac{1}{n}\sum_{i=1}^{n}|h_1(i) \oplus h_2(i)| \tag{5.4}$$

Algorithm B-DCT sign-based image hashing

The DCT sign only correlation (DSOC) function is applied to high-accuracy image registration proposes (Kuglin and Hines, 1975; Chen et al., 1994; Kuglin and Hines, 2002; Arnia et al., 2006), when dealing with image translation. Let us consider two images, $f(n_1,n_2)$ and $g(n_1,n_2)$. Let $F_C(k_1,k_2)$ and $G_C(k_1,k_2)$ denote the 2D-DCTs of the two images, respectively. In this case, the DCT coefficients are real numbers. The normalised cross spectrum $R_C(k_1,k_2)$ of those images is given by:

$$\begin{aligned} R_C(k_1,k_2) &= \frac{F_C(k_1.k_2)}{|F_C(k_1,k_2)|} \cdot \frac{G_C(k_1.k_2)}{|G_C(k_1,k_2)|} \\ &= sgn(F_C) \cdot sgn(G_C) \\ &= F_S(k_1,k_2) \cdot G_S(k_1,k_2) \end{aligned} \tag{5.5}$$

Where $F_S(k_1,k_2)$ and $G_S(k_1,k_2)$ are the signs values of $F_C(k_1,k_2)$ and $G_C(k_1,k_2)$, respectively. The DSOC function $r_c(n_1,n_2)$ is obtained by applying the inverse DCT (2D IDCT) on $R_C(k_1,k_2)$ as given by:

$$r_c(n_1,n_2) = IDCT(F_S(k_1,k_2) \cdot G_S(k_1,k_2)) \tag{5.6}$$

r_c is a 2D function which roughly represents the similarity between two images. Researchers have observed that the DSOC corresponds to a peak if one image is a shifted version of the other. This makes the magnitude of the peak invariant to shifting, although its location may vary depending on the amount of shifting. The height of the peak can be used as a good similarity measure for image matching. The height of the peak function is given by:

$$D = max(r_c(n_1,n_2)) \tag{5.7}$$

Figure 5.6 shows an example of image matching using the DSOC function. Where two images are similar, their DSOC function gives a distinct sharp peak. When two images are not similar, their peaks fall significantly. Thus, the DSOC function exhibits a much higher discrimination capability.

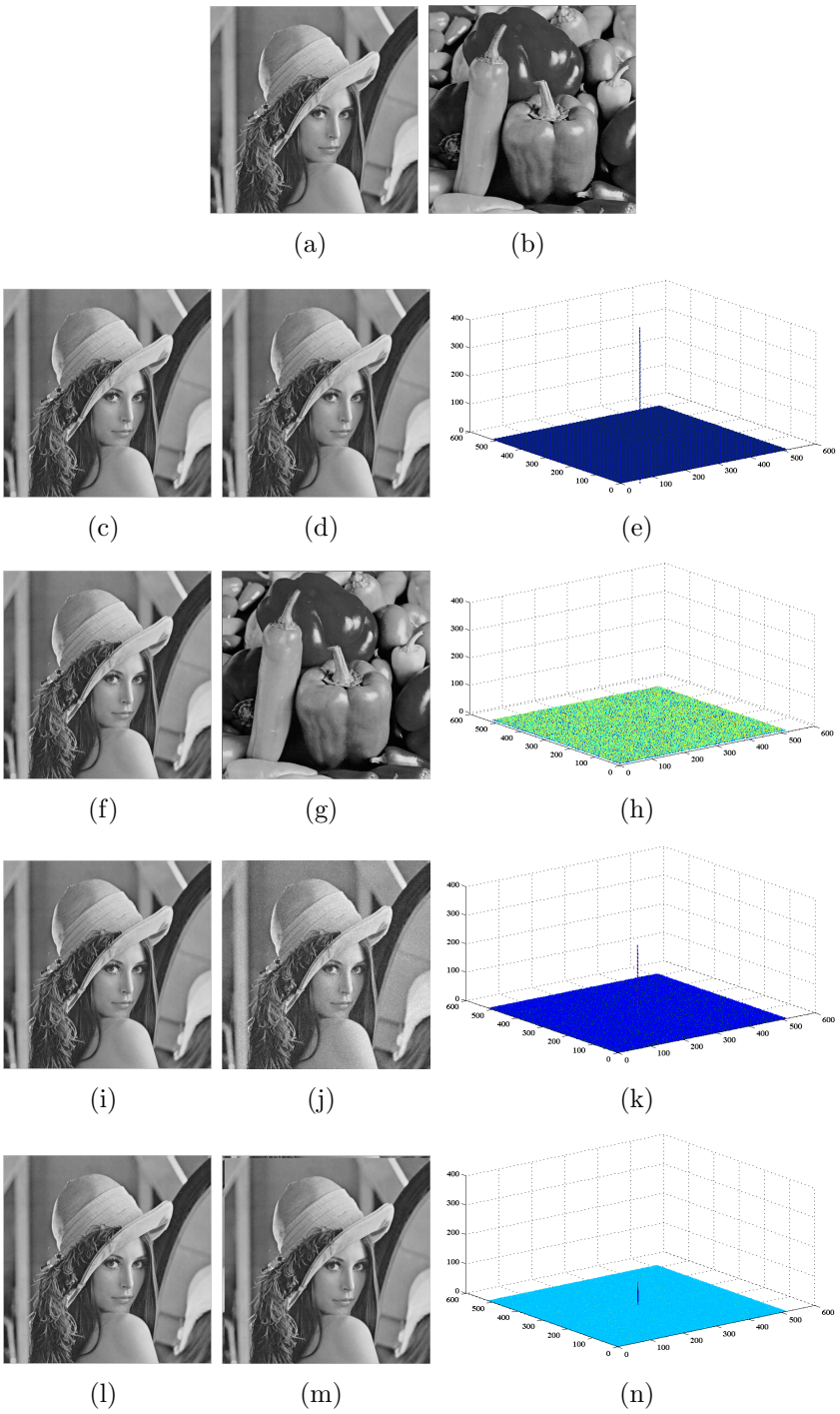


Figure 5.6: Example of DSOC function, (a) image "Lena" $f(n_1, n_2)$. (b) image "Peppers" $g(n_1, n_2)$. (e) DSOC function between the two original images image (a) and (b). (h) DSOC function between two different images (f) and (g). (k) DSOC function between the original image (i) and the noise image (j). (n) DSOC function between the original image and the shifted image (m).

5.2.2 Receiver Operating Characteristics analysis

To analyse the identification accuracy, ROC curves are also used to compare the performance of different image hashing techniques. To obtain ROC curves for the proposed image hashing algorithms, the TPR(τ) and FPR(τ) are defined as follows:
Algorithm A-DCT overlapping block-based image hashing:

$$TPR(\tau) = Probability(D(H(I,K), H(\hat{I},K)) < \tau) \tag{5.8}$$

$$FPR(\tau) = Probability(D(H(I,K), H(J,K)) < \tau) \tag{5.9}$$

Algorithm B-DCT sign-based image hashing:

$$TPR(\tau) = Probability(D(H(I,K), H(\hat{I},K)) > \tau) \tag{5.10}$$

$$FPR(\tau) = Probability(D(H(I,K), H(J,K)) > \tau) \tag{5.11}$$

Where τ is the identification threshold: The ROC curves are generated by sweeping the threshold τ from the minimum to the maximum value of all the distances between the manipulated and the original images, in order to compare the performances of the image hashing approaches.

5.2.3 Database and content-preserving operations

A dataset 120 original gray scale natural images is constructed including around 90 classic benchmark images, obtained by (CVG-URG, 2007). The motivation to construct such a dataset is to simulate possible quality distortions of digital images such as noise in broadcasting, transmission or different format changes.The details are given in Table 5.1.

Table 5.1: Content-preserving operations (CPOs) with various parameters

Manipulation type	Parameters
Image processing operations	
JPEG lossy compression	Quality Factor $QF = 90\sim10$
Additive White Gaussian Noise (AWGN)	Standard deviation $\sigma = 20\sim45$
Median filtering	window size $3\sim 11$
Histogram Equalization	/
Geometric distortions	
Rotation	degree $3°\sim14°$
Translation	window size $2\sim14$

5.3 Experimental results with the proposed image hashing technique in DCT domain

5.3.1 Robustness testing

Results of algorithm A-DCT overlapping block-based image hashing

The robustness of the algorithm A-DCT overlapping block-based image hashing is investigated with respect to the overlapping blocks and non-overlapping blocks for the statistical invariant extraction under content-preserving operations (CPOs). The overlapping blocks were formed to extract the feature vector from the original image and its attacked versions. As known, the statistics of blocks were approximately invariant under image processing attacks, however, the use of overlapping blocks can be justified by their robustness under geometric changes (Khelifi and Jiang, 2010). For the sake of illustration, Table 5.2 depicts results on the robustness against a number of geometric attacks measured by the normalised Hamming distance, with regards to the use of overlapping and non-overlapping blocks. It can be clearly seen that the feature coefficients obtained through non-overlapping blocks suffer more significant changes when compared to those derived from overlapping blocks. Additionally, the results of the robustness of the proposed algorithm A against image pro-

cessing operations and geometric attacks with respect to the size of overlapping blocks are plotted in Figures 5.7 to 5.11. It can be seen that the proposed technique exhibits good robustness against different attacks especially JPEG lossy compression and median filter attacks. Note that the performance improved as the size of blocks increased. This is attributed to the fact that the larger blocks offer better statistical invariance under image processing operations and geometric attacks. It is meant here by statistical invariance the robustness of statistical values, such as mean and standard deviation against different operations performed on the images. Therefore, a block of size 64×64 with horizontal and vertical overlapping by 16 pixels is adopted in the rest of this chapter.

Table 5.2: Normalisation Hamming distance between the feature vectors extracted from the original and its attacked versions Ov.:Overlapping blocks by sixteen pixels Non-Ov.: Non-overlapping blocks

Image	**Attacks**	**Block size 32**		**Block size 64**	
		Non-Ov.	Ov.	Non-Ov.	Ov.
Lena	Rotation 3°	0.214	0.196	0.140	0.140
	Rotation 5°	0.300	0.280	0.250	0.180
	Translation 3×3	0.128	0.111	0.140	0.060
	Translation 5×5	0.203	0.181	0.156	0.090
Peppers	Rotation 3°	0.214	0.189	0.140	0.135
	Rotation 5°	0.308	0.290	0.234	0.230
	Translation 3×3	0.109	0.106	0.093	0.070
	Translation 5×5	0.187	0.162	0.140	0.120
Baboon	Rotation 3°	0.218	0.193	0.144	0.140
	Rotation 5°	0.289	0.255	0.203	0.180
	Translation 5×5	0.132	0.105	0.078	0.070
	Translation 5×5	0.175	0.168	0.100	0.093

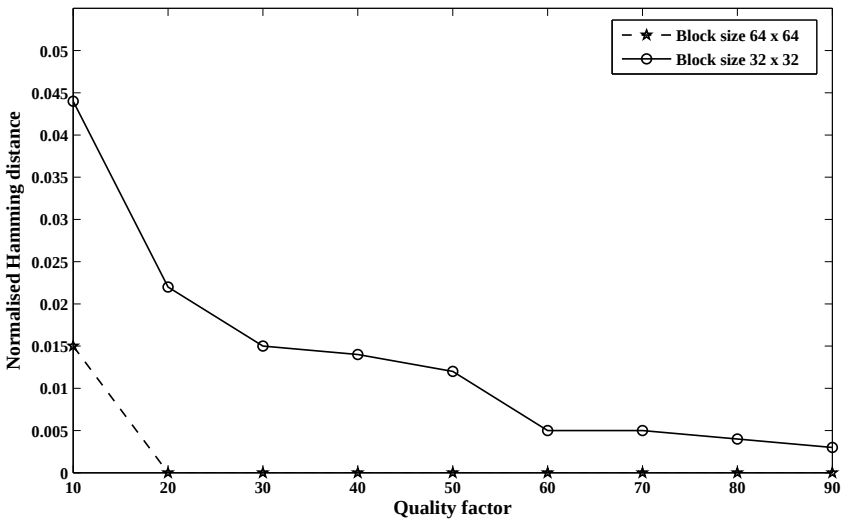


Figure 5.7: Performance of algorithm A using different block sizes for "Lena" image under JPEG lossy compression attack with 16 pixels overlap

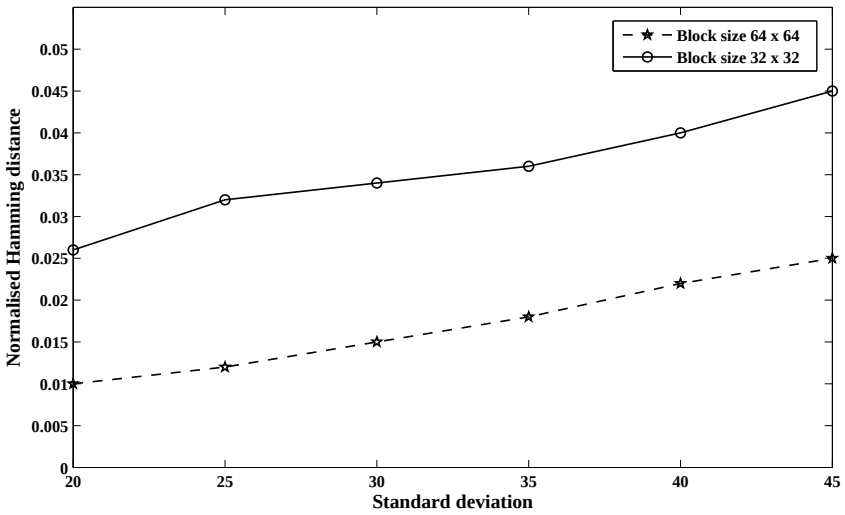


Figure 5.8: Performance of algorithm A using different block sizes for "Lena" image under additive white Gaussian noise attack with 16 pixels overlap

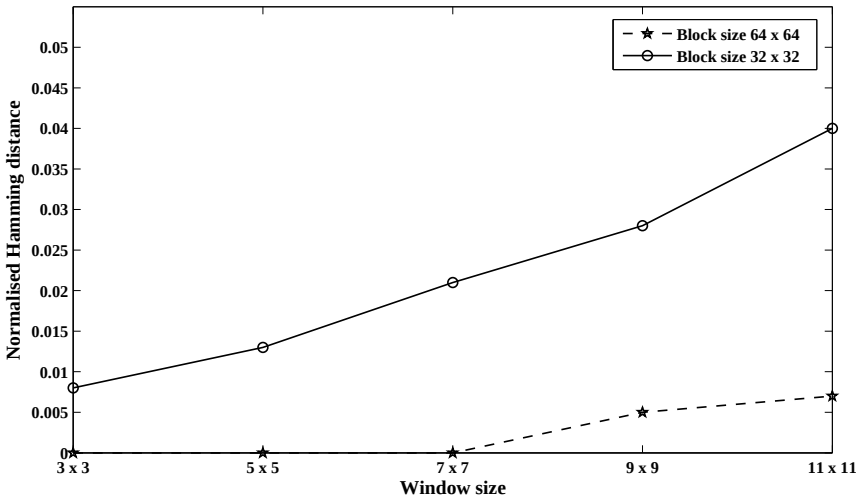


Figure 5.9: Performance of algorithm A using different block sizes for "Lena" image under median filtering attack with 16 pixels overlap

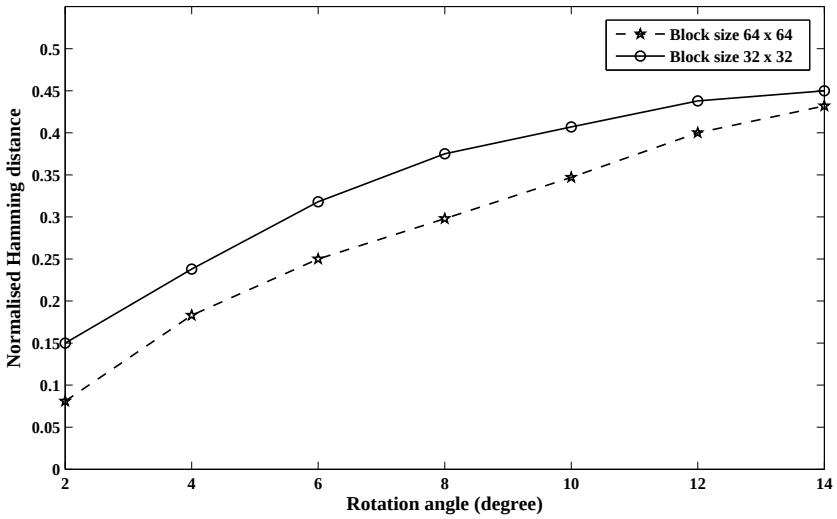


Figure 5.10: Performance of algorithm A using different block sizes for "Lena" image under rotation attack with 16 pixels overlap

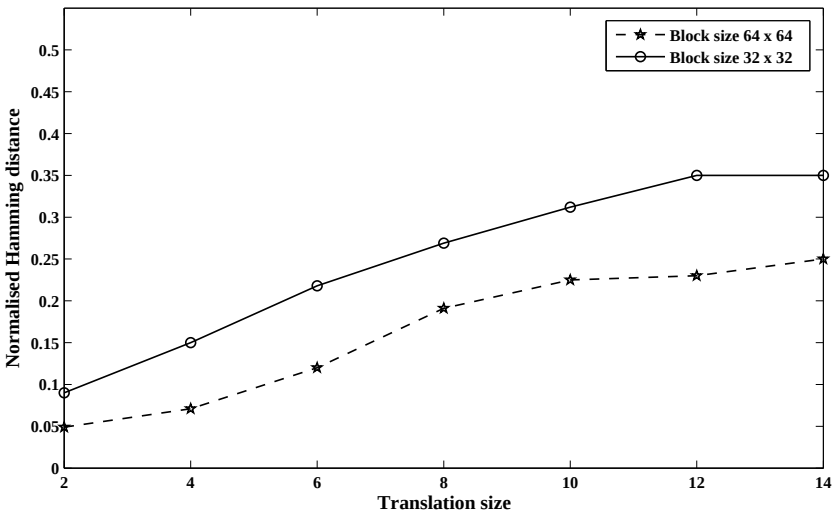


Figure 5.11: Performance of algorithm A using different block sizes for "Lena" image under translation attack with 16 pixels overlap

Results of algorithm B-DCT sign-based image hashing

The robustness of algorithm B-DCT sign-based image hashing algorithm has been assessed with respect to the size of the sign coefficients under six kinds of distortions as listed in Table 5.1. The selected hash length n were 512, 1024 and 2048 bits. The rate of the peak of DSOC for distorted images in relation to the peak of DSOC for original images is shown in Tables 5.3 to 5.5. From the results, it can be observed that the proposed algorithm B-DCT sign-based hashing is still superior under different image processing operations that include JPEG lossy compression, median filter, AWGN and histogram equalisation attacks and reveal the best performance from 89% to 100%. The results under the translation attack shows that the rate varies from 24% to 80% while the results under the rotation attack, vary from 7% to 57%. It was found that the choice of 512 sign coefficients yields a better performance under different types of attacks. This is attributed to that fact that the low frequency DCT coefficients are more stable than other frequency coefficients when attacked. Moreover, the DCT sign coefficients contained information that was related to the DFT phase. The phase of DFT is very important for image processing; even phase

information alone without the magnitude information of the image can be used to restore a significant part of the original version. In Table 5.3, some exceptions have been observed for additive noise attack with $\sigma = 20$ and histogram equalisation attack. This is due to the image content itself where some images show higher and hence more robust coefficients in the high frequency range than others. The tables shows the limitation of the DCT sign-based image hashing technique to deal with rotation attacks. In the rest of this chapter, DCT sign coefficients of size 512 bits are used.

Table 5.3: Peak of DSOC under different attacks in comparison with the original "Lena" image under different attacks

Image	**Attacks**	**512 bits**	**1024 bits**	**2048 bits**
Lena	Median filter 3×3	99.10%	98.09%	97.73%
	Median filter 5×5	98.60%	97.47%	95.98%
	Median filter 7×7	97.27%	95.10%	93.25%
	JPEG compression QF 90	100%	100%	100%
	JPEG compression QF 50	100%	99.83%	98.89%
	JPEG compression QF 10	97.24%	94.82%	92.20%
	AWGN σ=20	97.79%	97.94%	96.52%
	AWGN σ=25	97.06%	96.11%	95.14%
	AWGN σ=30	96.93%	96.14%	95.42%
	Histogram Equalisation	91.67%	90.52%	90.92%
	Rotation 3°	49.47%	41.77%	28.03%
	Rotation 5°	32.11%	23.07%	11.73%
	Rotation 7°	21.41%	11.69%	7.67%
	Translation 3×3	78.55%	70.31%	60.61%
	Translation 5×5	66.34%	54.54%	40.00%
	Translation 7×7	56.45%	40.25%	24.97%

Table 5.4: Peak of DSOC under different attacks in comparison with the original "Peppers" image under different attacks

Image	**Attacks**	**512 bits**	**1024 bits**	**2048 bits**
Peppers	Median filter 3×3	99.93%	99.18%	99.15%
	Median filter 5×5	97.88%	97.20%	96.98%
	Median filter 7×7	96.25%	96.10%	94.96%
	JPEG compression QF=90	99.65%	98.76%	99.85%
	JPEG compression QF=50	99.12%	99.20%	98.85%
	JPEG compression QF=10	98.85%	97.82%	95.82%
	AWGN σ=20	99.87%	98.76%	97.56%
	AWGN σ=25	99.17%	98.47%	96.80%
	AWGN σ=30	98.44%	98.08%	95.75%
	Histogram Equalisation	94.46%	93.61%	92.65%
	Rotation 3°	55.63%	41.83%	29.23%
	Rotation 5°	36.90%	21.80%	12.78%
	Rotation 7°	25.02%	16.59%	8.92%
	Translation 3×3	79.98%	74.99%	62.57%
	Translation 5×5	69.15%	61.38%	42.59%
	Translation 7×7	56.15%	43.22%	26.08%

Table 5.5: Peak of DSOC under different attacks in comparison with the original "Baboon" image under different attacks

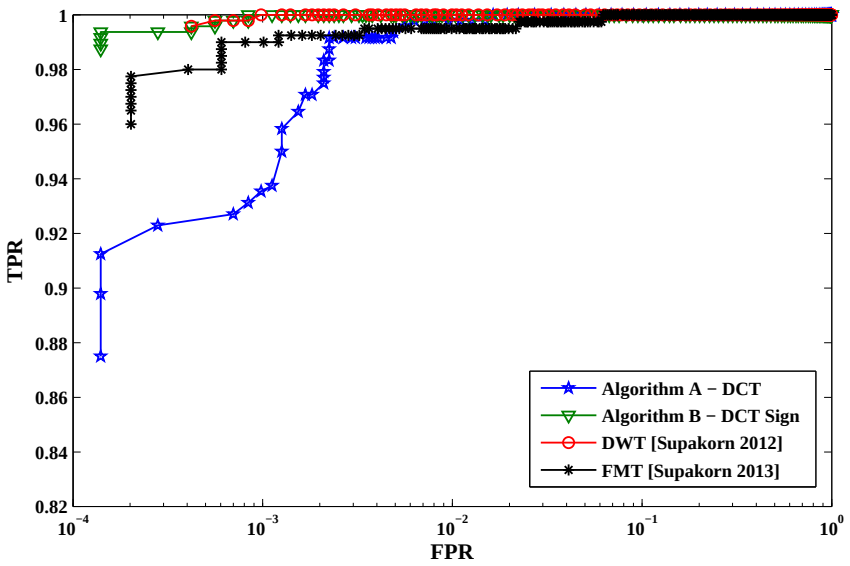
Image	Attacks	512 bits	1024 bits	2048 bits
Baboon	Median filter 3×3	97.53%	96.19%	95.39%
	Median filter 5×5	96.39%	93.63%	92.67%
	Median filter 7×7	93.59%	90.35%	89.57%
	JPEG compression QF=90	99.99%	99.64%	99.36%
	JPEG compression QF=50	99.36%	98.28%	98.22%
	JPEG compression QF=10	96.34%	92.79%	92.05%
	AWGN σ=20	99.64%	98.08%	96.44%
	AWGN σ=25	98.38%	96.83%	94.61%
	AWGN σ=30	97.99%	96.83%	93.86%
	Histogram Equalisation	91.15%	90.61%	89.51%
	Rotation 3°	57.42%	49.40%	33.47%
	Rotation 5°	44.67%	34.27%	19.26%
	Rotation 7°	35.34%	23.95%	16.12%
	Translation 3×3	80.50%	73.80%	64.20%
	Translation 5×5	69.99%	58.35%	41.61%
	Translation 7×7	59.93%	43.55%	24.16%

5.3.2 Robustness versus discriminability

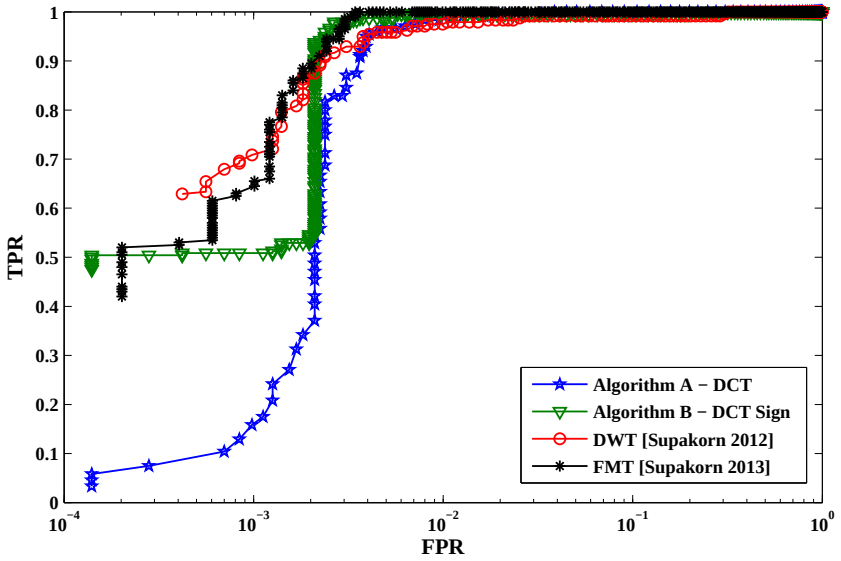
In this section, a statistical comparison of the proposed algorithm A-DCT overlapping block-based image hashing, algorithm B-DCT sign-based image hashing, sub-images-DWT-based image hashing and FMT-based image hashing algorithms has been performed by studying the corresponding ROC curves. It is worth pointing outing that the distance used to compute TPR and FPR depends on the nature of the hash. The normalised Hamming distance (NHD) is used for binary hashes extracted via sub-images-DWT-based image hashing and DCT overlapping block-based image hashing. The Euclidian distance is used for the real-value hash extracted via the FMT-based image hashing technique. The peak of the DSOC is used as a similarity measure between two hash vectors extracted via DCT sign-based image hashing.

Firstly, the overall ROC curves for all types of manipulations when using different image hashing techniques are generated, and the resulting ROC curves are shown in Fig-

ure 5.12. Figure 5.12 (a), shows that the algorithm B-DCT sign-based image hashing and sub-images-DWT-based image hashing techniques performed equally well under image processing operations closely followed by algorithm A-DCT overlapping block-based image hashing and the FMT-based image hashing techniques. Figure 5.12 (b), shows that the FMT-based image hashing technique provides the best performance under geometric attacks in the range of TPR > 0.979, FPR $> 10^{-2.519}$ closely followed by the B-DCT sign-based hashing image technique which becomes the most efficient in the range TPR > 0.901, FPR $> 10^{-2.678}$. While sub-images-DWT-based image hashing and algorithm A-DCT overlapping block-based image hashing techniques are slightly sensitive. Moreover, in order to assess robustness to each type of attacks separately and obtain a clear view, an ROC curve is also generated for each particular attack and hash algorithm as shown in Figures 5.13 to 5.15. It can be seen that the proposed algorithm B-DCT sign-based image hashing technique offers the best performance, compared to others (see Figure 5.13 and Figure 5.14 (a)). This is because the signatures generated from the corresponding coefficients are invariant against such attacks. As can be seen the proposed DCT sign-based image hashing technique performs extremely well under the translation attack as shown in Figure 5.15 (b), and gives a comparable performance with FMT-based image hashing under the rotation attack in the range of TPR > 0.958, FPR $> 10^{-2.518}$. Beyond the range, the performance of all the techniques tends to drop significantly (see Figure 5.15 (a)).

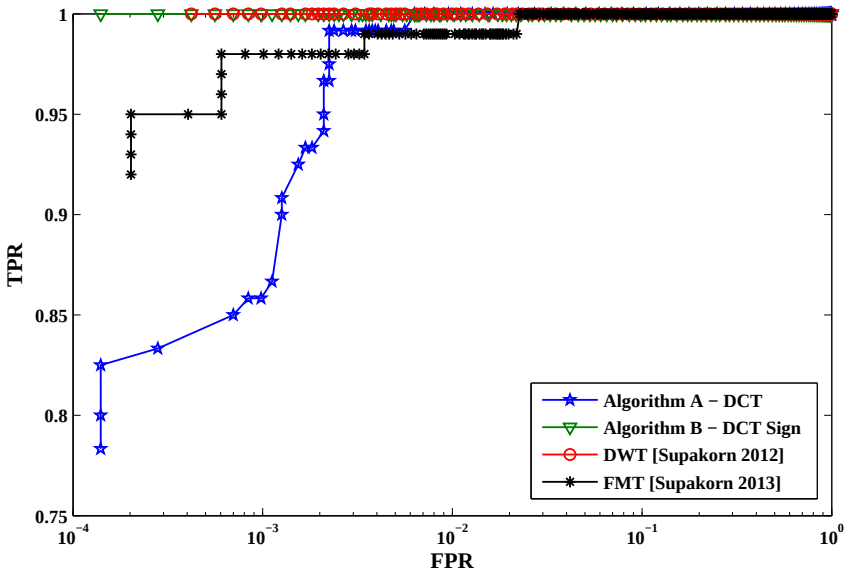


(a) Image processing operations

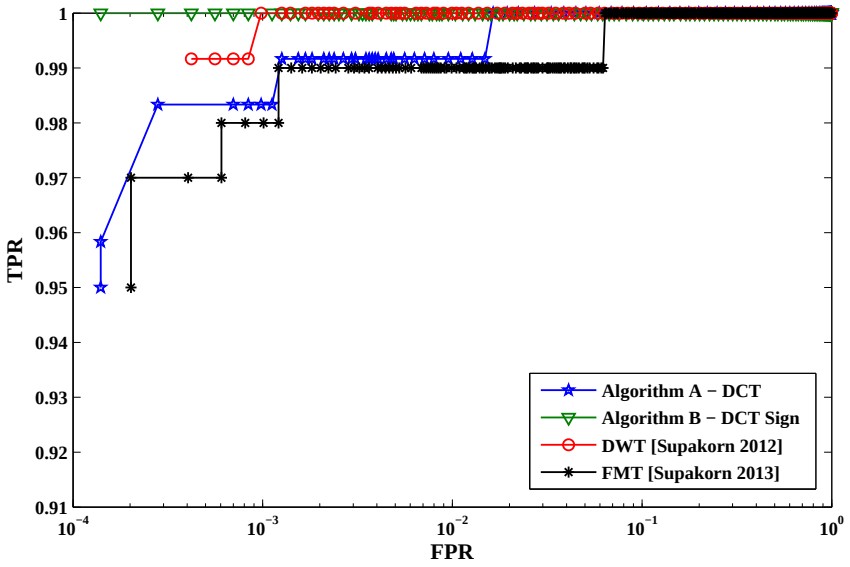


(b) Geometric attacks

Figure 5.12: The overall ROC curves for all types of test manipulations when applying different hashing schemes, (a) Image processing operations (b) Geometric attacks

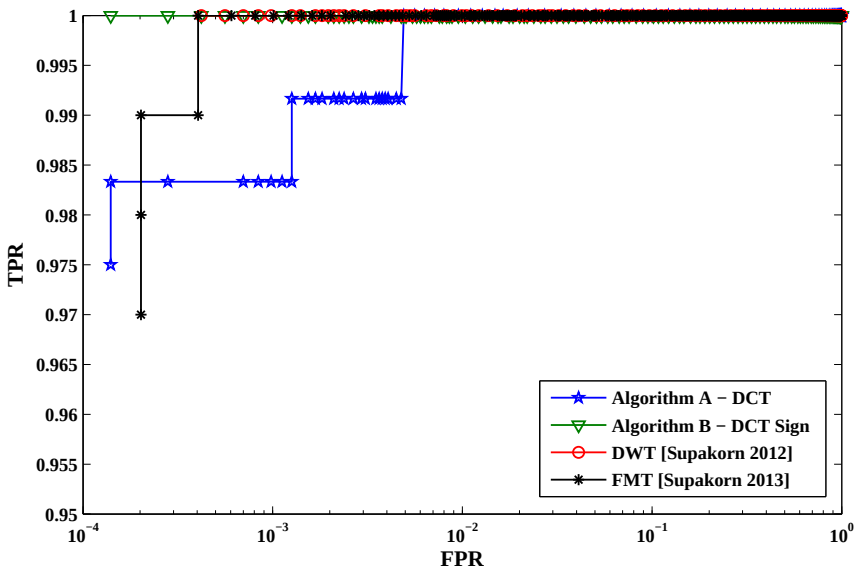


(a) JPEG lossy compression

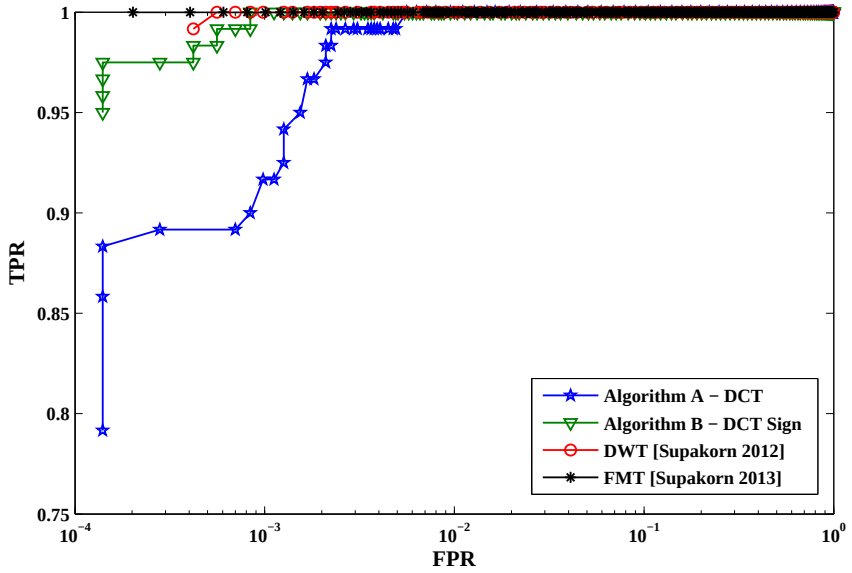


(b) Median filtering

Figure 5.13: ROC curves for each type of manipulation when applying it to different image hashing schemes, (a) JPEG lossy compression (b) Median filtering

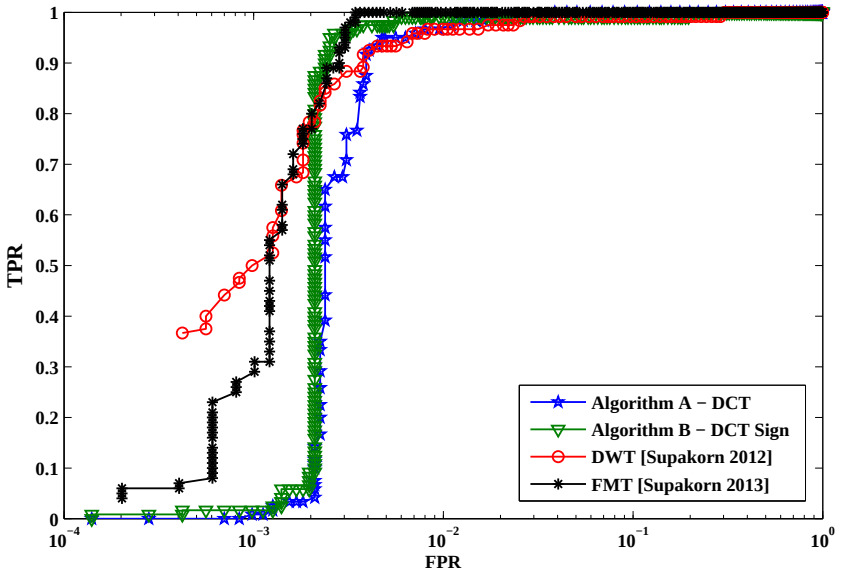


(a) AWGN

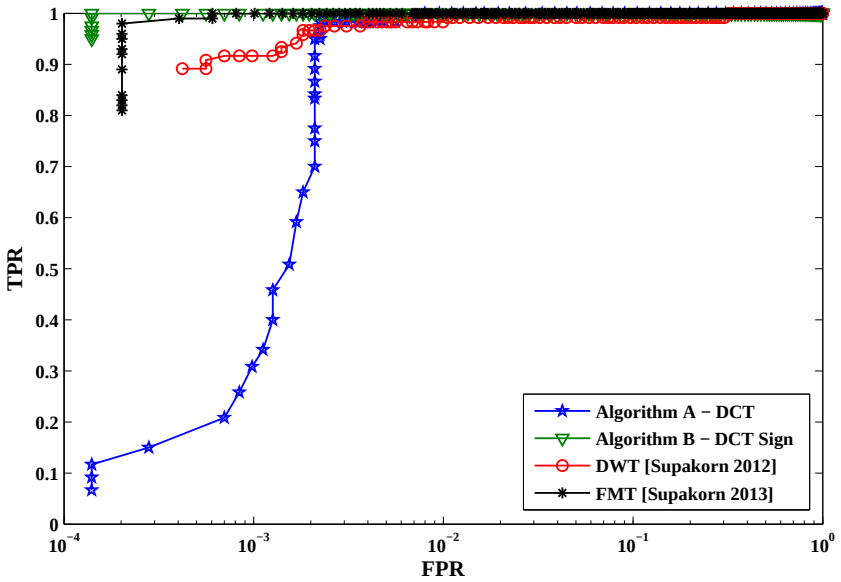


(b) Histogram equalisation

Figure 5.14: ROC curves for each type of manipulation when applying it to different image hashing schemes, (a) AWGN (b) Histogram equalisation



(a) Rotation



(b) Translation

Figure 5.15: ROC curves for each type of manipulation when applying it to different image hashing schemes, (a) Rotation (b) Translation

5.3.3 Unpredictability testing

The security of image hashing is another important property for the proposed image hashing techniques. Here, the security of image hashing is defined in terms of unpredictability of the hash and it is the main focus. The binomial distribution is used to estimate the correlations between different hashes. The actual distribution of 7140 normalised Hamming distance between the hash pairs of 120 different images in database as mentioned in section 5.2.3 is shown in Figure 5.16 and Figure 5.17, respectively. The empirical distribution of the proposed DCT overlapping block-based image hashing and DCT sign-based image hashing is displayed in Table 5.6. Once again, Table 5.6, DCT block-based image hashing shows a standard deviation of $\sigma = 0.0603$, with a mean of $\mu = 0.494$, a binomial process with $N = 70$. The likelihood of two binary hashes from different images matching completely by chance is one in 2^{70}, or approximate 10^{-21}. This indicated that 70 out of 320 hash bits (22%) were independent and unpredictable. The DCT sign-based image hashing has a standard deviation of $\sigma = 0.0250$, with a mean of $\mu = 0.243$. A binomial process with $N = 400$. by chance is one in 2^{400}, or approximately 3×10^{-120}. This means that, 400 out of 512 hash bits (78%) were independent and unpredictable. Obviously, DCT sign-based image hashing is more secure than DCT overlapping block-based image hashing. This is attributed to the fact that the information extracted from overlapping blocks is redundant. This suggests that a number of hash bits share the same information about the image content.

Table 5.6: The empirical distribution of hash values Ov.:Overlapping blocks

Proposed	**Stand deviation**	**mean**	**N**	**independent and unpredictable**
DCT Ov.-based	0.0603	0.494	70	70 out of 320 hash bits (22%)
DCT sign-based	0.0250	0.243	400	400 out of 512 hash bits (78%)

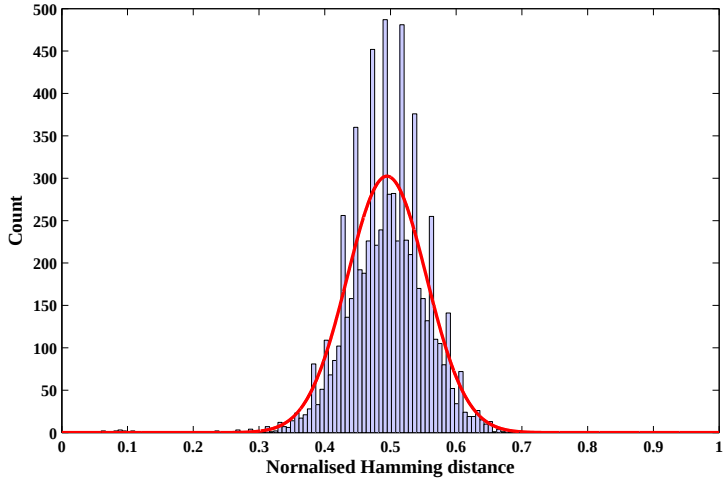


Figure 5.16: Distribution normalised Hamming distance between distinct hashes of algorithm A-DCT overlapping block-based

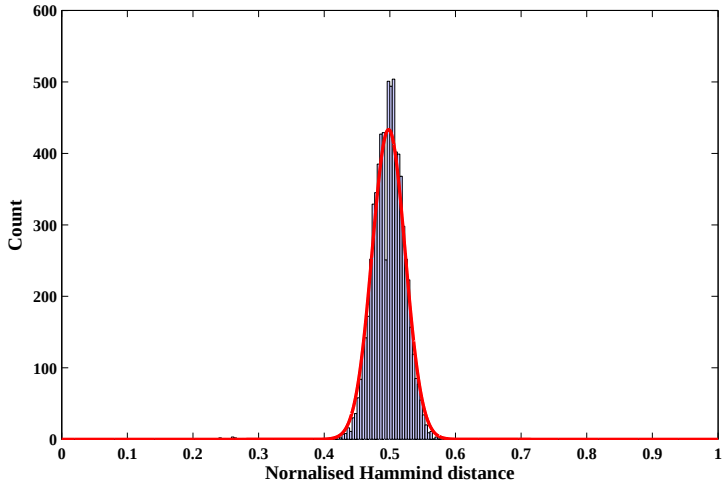


Figure 5.17: Distribution normalised Hamming distance between distinct hashes of algorithm B-DCT sign-based

5.4 Summary

In this chapter, two new techniques of perceptual image hashing in DCT domain, DCT overlapping block-based image hashing and DCT sign-based image hashing techniques are

presented. The main idea was to extract robust features in the transform domain and obtain a hash. Both image hashing techniques were investigated against a large class of content-preserving operations (CPOs). Furthermore the techniques were compared with related state-of-the art image hashing algorithms and shown to perform well under image processing operations and geometric attacks. From the experimental results, it was observed that the low frequency coefficients for DCT sign-based image hashing were robust to a large class of content-preserving operations (CPOs). Compared with the DCT overlapping block-based image hashing, FMT-based image hashing and sub-images-DWT-based image hashing techniques, the DCT sign-based image hashing technique, which exploits the desirable property of representing the image with a few sign coefficients, offers an enhanced performance. However, the inherent drawback of the DCT-sign approach is not invariant against rotation attacks.

Chapter 6

Conclusion

6.1 Contribution of the thesis.

In this thesis, perceptual image hashing has been investigated, from theory to applications. This thesis has presented new techniques for improving the perceptual robustness and security of image hashing, proposed and evaluated on image dataset. We categories the image hashing algorithm according to its major components, including pre-processing of images, feature extraction and post processing. For security purpose, image hashing occurring from pseudo-randomisation was analysed. Furthermore, a number of state-of-the art image hashing techniques have been reviewed. We easily obtained the perspectives for the recent developments in area of research in image hashing. A considerable amount of the work investigated the robustness against content-preserving operations (CPOs) such as additive white Gaussian noise, JPEG lossy compression, median filtering, rotation and translation attacks. These types of distortions are the most common ones, to which the proposed scheme should be robust enough to achieve reliability. Researchers in the area of perceptual image hashing should consider their algorithms to deal with these preliminary distortions.

The first investigation of the thesis is in the discrete wavelet transform (DWT) domain, which applies pseudo-random sub-images and a recent dimension reduction technique, re-

ferred to as non-negative matrix factorisation (NMF), into the hash generation. The image is divided into overlapping sub-images by pseudo-random sampling and treated as image features, which could be further a feature extracted by DWT. The image is then projected into a lower dimensional space by NMF in order to generate a compact image hash. Finally, the hash vector is generated. The pseudo-random sub-image and NMF are desirable properties to generate robust and secure image hashes. From the robustness of the proposed image hashing demonstrated in experiments, the image hashing technique was investigated with regard to the use of the size of sub-image and number of sub-images. From the experimental results, it was noted that the pseudo-random sub-images and NMF can be incorporated into DWT image hashing to improve the performance under content-preserving operations (CPOs), such as JPEG lossy compression and median filter etc., but they were sensitive to geometric attacks.

In the second investigation of the project, a novel image hashing algorithm based on the Fourier-Mellin transform (FMT) domain was presented. FMT was successfully used in many image registration applications, image hashing and image watermarking areas. The proposed FMT-based image hashing scheme was shown to be more robust than the sub-images-DWT-based image hashing under geometric attacks. The major benefit of employing FMT is its spectrum is invariant to rotation, translation and scaling. However, the inherent drawback of the Fourier transform makes FMT only robust to geometric transform, but vulnerable to other image processing operations such as cropping and noise. This is due to the reason that when an image is converted into the spectrum domain by 2D FT, each coefficient value involves all the pixels of the image. It means that the Fourier coefficients are dependent on the global information of the image in the spatial domain. Therefore, we incorporated the low pass filtering and histogram equalisation into the proposed FMT image hashing scheme to improve its performances under image processing operations. Based on the experimental results, it was shown that FMT-based image hashing gives a superior robustness against various content-preserving operations (CPOs), especially translation attacks. The limitation of the FMT-based image hashing technique is discriminability. Discriminability suggests that two different images should provide two

dissimilar hashes. The FMT used the Fourier spectrum which fails to differentiate between images of different contents.

The third investigation of the project is perceptual image hashing based on the discrete cosine transform (DCT). As known, the DCT has been used in image compression standards. The DCT coefficients exhibit characteristics of the image that were able to survive distortions e.g.a format change, or re-compression. The significance of the proposed DCT-based image hashing lies in two aspects: algorithm A-DCT overlapping block-based image hashing and algorithm B-DCT sign-based image hashing. The DCT overlapping block-based image hashing technique used the low frequency coefficients and then used a randomisation technique to enhance security. The image is divided into overlapping blocks with a particular size and then a small number of low frequency coefficients are selected from the DCT block. Form the experiments, the performance of the proposed algorithm A-DCT overlapping block-based image hashing performed well under image processing attacks, although it remained sensitive to geometric attacks. In algorithm B-DCT sign-based image hashing, the main idea was to use the energy compaction property of the DCT sign values, which carry a significant part of information, representing the image in terms of texture, contours and areas of edges, as perceptual features. The proposed B-DCT sign-based image hashing scheme was demonstrated to be more robust than the proposed A-DCT overlapping block-based image hashing under geometric attacks. In relation to the experimental results, the scheme B-DCT sign-based image hashing produced a superior performance under content-preserving operations (CPOs). The DCT sign-based image hashing scheme has also been shown to be the most secure technique compared to other techniques proposed in this research as it offers the highest rate of bit independence in a hash. However, the drawback of proposed B-DCT sign-based image hashing technique resides in its limited robustness against rotation attacks.

To the best of the author's knowledge, no existing work provides a complete solution for perceptual image hashing with respect to content identification issues, and these approaches do have limitations. However, the proposed image hashing techniques reported in this thesis provide promising results and tackling important issues of image hashing such as

robustness and discriminability. With the result obtained throughout the thesis, the aims and objectives set out in section 1.3 have been satisfactorily achieved. Nevertheless, with technology development in this area, more work is required to ensure a trade-off between robustness and security in perceptual image hashing.

6.2 Recommendation and Future work.

This thesis has presented a novel image hashing algorithm in different transform domain. Although promising results have been satisfactorily achieved, we would like to underline some future research directions at the end of this thesis.

- The DCT-sign based image hashing scheme proposed in chapter 5 has shown very good performance in terms of robustness and discriminability under different signal processing attacks. It has also been shown to be robust against translation attacks. However, the results showed that it still suffers from small rotations because the DSOC function can recover translations only. Therefore, as part of our future work plan, we propose to apply a log-polar transform to the image prior to applying the DCT and extracting DCT-sign coefficients that construct the hash. The log-polar transform converts any rotational changes into vertical shifting with a proportional amount of translation (see Figure 2.2).

- The security of the proposed hashing techniques has been assessed in terms of the unpredictability. This gives the proportion of independent bits in a hash. However, another security measure, called unicity distance, consists of determining the number of image/hash pairs required to estimate the secret key. This measure will be used in future to analyse further the security of the proposed systems.

- Visually salient regions play an essential role in the determination of video copies because most of the attacks that occur on video data tend to keep the significant parts of the video in order to maintain its perceptual content and alter non-salient regions. In order to exploit this feature, a judicious idea would be to design a

hashing algorithm that assigns high weights to hash bits corresponding to salient regions during the hash matching process. This way, salient regions-based hashing can be used to identity image content because the attacker cannot destroy the salient regions from which the hash is extracted.

- Perceptual hashing will be used as an image-dependent key to secure watermarking systems. Indeed, one of the main weakness of watermarking systems resides in the use of a constant key for large number of images. The multi-use of the same key allows the attacker to estimate the key and break the security of the system.

Bibliography

Alghoniemy M.and Tewfik, A. H. (2004), 'Geometric invariance in image watermarking', *Image Processing, IEEE Transactions on* **13**(2), 145–153.

Antonini, M., Barlaud, M., Mathieu, P. and Daubechies, I. (1992), 'Image coding using wavelet transform', *Image Processing, IEEE Transactions on* **1**(2), 205–220.

Arnia, F., Fujiyoshi, M. and Kiya, H. (2007), The use of DCT coefficient sign for content-based copy detection, *in* 'Communications and Information Technologies, 2007. ISCIT '07. International Symposium on', pp. 1476–1481.

Arnia, F., Iizuka, I., Fujiyoshi, M. and Kiya, H. (2007*a*), 'Dct Sign-Based Similarity Measure for JPEG Image Retrieval', **E90-A**(9), 1976–1985.

Arnia, F., Iizuka, I., Fujiyoshi, M. and Kiya, H. (2007*b*), Fast Method for Joint Retrieval and Identification of JPEG Coded Images Based on DCT Sign, *in* 'Image Processing, 2007. ICIP 2007. IEEE International Conference on', Vol. 2, pp. II – 229–II – 232.

Arnia, F., Iizuka, I., Kobayashi, H., Masaaki, F. and Kiya, H. (2006), Dct Sign Only Correlation and Its application to Image Registration, *in* 'Circuits and Systems, 2006. APCCAS 2006. IEEE Asia Pacific Conference on', pp. 466–469.

Arnia, F., Munadi, K., Fujiyoshi, M. and Kiya, H. (2009), Efficient content-based copy detection using signs of DCT coefficient, *in* 'Industrial Electronics Applications, 2009. ISIEA 2009. IEEE Symposium on', Vol. 1, pp. 494–499.

Blahut, R. E. (1983), Theory and Practice of Error Control Codes, Addison-Wesley Publishing Company.

Bracewell, R. (1999), *The Fourier transform and its applications,Third edition*, McGraw-Hill,New York, USA.

Brasnett, P. and Bober, M. (2008), Fast and robust image identification, *in* 'Pattern Recognition, 2008. ICPR 2008. 19th International Conference on', pp. 1–5.

Burrus, C. S., Gopinath, R. A. and Guo, H. (1998), *Introduction to wavelets and wavelet transforms*, Prentice-Hall, New Jersey.

Chen, Q. S., Defrise, M. and Deconinck, F. (1994), 'Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition', **16**(12), 1156–1168.

Cox, I. J., Kilian, J., Leighton, T. and Shamoon, T. (1996), Secure spread spectrum watermarking for images, audio and video, *in* 'IEEE Trans. on Image Processing', Vol. 6, pp. 243–246.

Cox, I. J., Miller, M. L. and Bloom, J. A. (2000), Watermarking applications and their properties, *in* 'Information Technology: Coding and Computing, 2000. Proceedings International Conference on', pp. 6–10.

CVG-URG (2007), 'Test Image'. Accessed: 29 July 2013.
URL: *http://decsai.ugr.es/cvg/dbimagenes/*

Daugman, J. G. (1993), 'High confidence visual recognition of persons by a test of statistical independence', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **15**(11), 1148–1161.

Dittmann, J., Steinmetz, A. and Steinmetz, R. (1999), Content-based digital signature for motion pictures authentication and content-fragile watermarking, *in* 'Multimedia Computing and Systems, IEEE International Conference on', Vol. 2, pp. 209–213 vol.2.

Egan, J. P. (1975), signal detection theory and ROC analysis, *in* 'Series in Cognition and Perception, Academic Press'.

Fawad, A. and Siyal, M. (2005), A Secure and Robust Hashing Scheme for Image Authentication, *in* 'Information, Communications and Signal Processing, 2005 Fifth International Conference on', pp. 705–709.

Fawad, A. and Siyal, M. (2006), A secure and Robust Wavelet-Based Hashing Scheme for Image Authentication, Vol. 4352, pp. 51–62.

Fawad, A., Siyal, M. Y. and Vali, U. A. (2010), 'A secure and robust hash-based scheme for image authentication', *Signal Processing* **90**(5), 1456–1470.

Fawcett, T. (2006), 'An introduction to ROC analysis', *Pattern Recognition Letters* **27**(8), 861 – 874.

Fodor, I. K. (2002), A survey of dimension reduction techniques, Technical report, Tech. Rep., US DOE Office of Scientific and Technical Information.

Fridrich, J. (2000), 'Visual hash for oblivious watermarking', *Proc. SPIE* **3971**, 286–294.

Fridrich, J. and Goljan, M. (2000), Robust hash functions for digital watermarking, *in* 'Proc.IEEE Int.Conf.on Information Technology:Coding and computing', pp. 178–183.

Gerold, L. and Andreas, U. (2008), 'Key-dependent JPEG2000-based robust hashing for secure image authentication', *EURASIP J. Inf. Secur.* **2008**, 1:1–1:19.

Guo, X. and Hatzinakos, D. (2007), Content based image hashing via wavelet and radon transform, *in* 'Advances in Multimedia Information Processing PCM 2007', Vol. 4810 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 755–764.

Guo, X., Xu, Z., Lu, Y. and Pang, Y. (2005), An application of fourier-mellin transform in image registration, *in* 'Computer and Information Technology, 2005. CIT 2005. The Fifth International Conference on', pp. 619–623.

Hadmi, A., Puech, W., Said, B. A. E. and Ouahman, A. A. (2010), Analysis of the robustness of wavelet-based perceptual signatures, *in* 'Image Processing Theory Tools and Applications (IPTA), 2010 2nd International Conference on', pp. 112–117.

Hampapur, A. and Bolle, R. M. (2001), Comparison of distance measures for video copy detection, *in* 'Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on', pp. 737–740.

Harmanci, O., Monga, V. and Mihcak, M. K. (2006), Geometrically invariant image watermarking via robust perceptual hashes, *in* 'Image Processing, 2006 IEEE International Conference on', pp. 1397–1400.

Hassan, M. A., Hasan, M. Y. Y. and Wahab, A. A. M. (2012), 'Robust Visual Hashing for Image Authentication', pp. 763–767.

Hernandez, R. A. P., Miyatake, M. N. and Kurkoski, B. M. (2011), Robust image hashing using image normalization and SVD decomposition, *in* 'Circuits and Systems (MWSCAS), 2011 IEEE 54th International Midwest Symposium on', pp. 1–4.

Hu, Y. and Niu, X. (2010), Dwt based robust image hashing algorithm, *in* 'Networked Computing (INC), 2010 6th International Conference on', pp. 1–4.

International Standard, ISO/IEC/JTC1/SC29 WG11 (1998), *ISO/IEC 13818-3, Information technology – generic coding of moving pictures and associated audio information – Part 3: Audio*, International Organization for Standardization, Geneva, Switzerland.

Jain, A. K., Farelle, P. M. and Algazi, V. R. (1984), Digital image processing techniques, *in* 'Digital Image Processing Techniques, M. P. Ekstrom Editor', Academic Press, pp. 171–226.

Jie, Z. (2013), A Novel Block-DCT and PCA Based Image Perceptual Hashing Algorithm, *in* 'Computing Research Repository (CoRR)'.

Kailasanathan, C., Naini, R. S. and Ogunbona, P. (2003), Compression Tolerant DCT Based Image Hash, *in* 'Proceedings of the 23rd International Conference on Distributed Computing Systems', ICDCSW '03, IEEE Computer Society, Washington, DC, USA, pp. 562–567.
URL: *http://dl.acm.org/citation.cfm?id=839280.840577*

Kailasanathan, C. and Nani, R. S. (2001), Image Authentication Surviving Acceptable Modofications Using Statistical Measures and K-mean Segmentation, *in* 'IEEE-EURASIP work. Nonlinear Sig. and Image Processing', Vol. 1, pp. 1–13.

Khelifi, F. and Jiang, J. (2010), Perceptual Image Hashing Based on Virtual Watermark Detection, *in* 'IEEE TRANSACTIONS ON IMAGE PROCESSING', Vol. 19, pp. 981–993.

Kim, C. (2003), 'Content-based image copy detection', *Signal Processing: Image Communication* **18**(3), 169 – 184.
URL: *http://www.sciencedirect.com/science/article/pii/S0923596502001303*

Kitanovski, V., Taskovski, D. and Bogdanova, S. (2007), 'Combined hashing/watermarking method for image authentication', **1**(6), 541 – 548.

Kondo, H. (2001), Application of DCT Sign Signal for Human face Recognition, *in* 'Proc. IEEE.M2VIPO1', Hong Kong.

Korattikara, A., Boyles, L., Welling, M., Kim, J. and Park, H. (2011), 'Journal of machine learning research - proceedings track', **15**, 128–136.

Kozat, S. S., Venkatesan, R. and Mihçak, M. K. (2004), Robust perceptual image hashing via matrix invariants, *in* 'Proceedings IEEE International Conference on Image Processing', Vol. 5, Singapore, pp. 3443–3446.

Kuglin, D. C. and Hines, C. D. (1975), The phase correlation image aligument method, *in* 'Proc. Int. Conf. on Cybernetics and Society', pp. 163–165.

Kuglin, D. C. and Hines, C. D. (2002), Phase-based image matching and its application to intelligent vision systems, *in* 'Proc. Int. Symp. New Paradign VLSI Computing', pp. 95–100.

Lee, D. D. and Seung, S. H. (1999), Learning the parts of objects by non-negative matrix factorization, *in* 'Nature 401', pp. 788–791.

Lee, D. D. and Seung, S. H. (2001), Algorithm for Non-negative Matrix Factorisation, *in* 'Proc. Advances in Neural Information Processing Systems Conference 9', Vancouver, BC, Canada, pp. 556–562.

Lefbvre, F., Macq, B. and Legat, J. D. (2002), 'RASH: Radon Soft Hash algorithm'.

Lefebvre, F., Czyz, J. and Macq, B. (2003), A robust soft hash algorithm for digital image signature, *in* 'Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on', Vol. 2, pp. II–495–8 vol.3.

Lei, Y., Wang, Y. and Huang, J. (2011), Robust image hash in Radon transform domain for authentication, *in* 'Signal Processing: Image Communication', Vol. 26, pp. 280 – 288. **URL:** *http://www.sciencedirect.com/science/article/pii/S0923596511000452*

Li, X. and Fukui, K. (2007), Fisher Non-negative Matrix Factorization with Pairwise Weighting., pp. 380–383.

Lin, C. Y. and Chang, S. F. (1998), 'Generting Robust Digital Signature for image/video Authentication'.

Lin, C. Y. and Chang, S. F. (2001), A robust image authentication method distinguishing JPEG compression from malicious manipulation, *in* 'Circuits and Systems for Video Technology, IEEE Transactions on', Vol. 11, pp. 153–168.

Lin, Y. C., Varodayan, D. and Girod, B. (2007), Image Authentication Based on Distributed Source Coding, *in* 'Image Processing, 2007. ICIP 2007. IEEE International Conference on', Vol. 3, pp. 5 – 8.

Lin, Y. C., Wu, M., Bloom, J. A., Cox, I. J., Miller, M. L. and Lui, Y. M. (2001), 'Rotation, scale, and translation resilient watermarking for images', *Image Processing, IEEE Transactions on* **10**(5), 767–782.

Lowe, D. G. (2004), Distinctive image features from scale-invariant keypoints, *in* 'International Journal of Computer Vision', Vol. 60, pp. 91–110.

Lu, C. S. and Hsu, C. Y. (2005), 'Geometric distortion-resilient image hashing scheme and its applications on copy detection and authentication', *Multimedia Systems* **11**(2), 159–173.
URL: *http://dx.doi.org/10.1007/s00530-005-0199-y*

Lu, C. S., Hsu, C. Y., Sun, S. W. and Chang, P. C. (2004), Robust mesh-based hashing for copy detection and tracing of images, *in* 'Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on', Vol. 1, pp. 731–734 Vol.1.

Lu, C. S. and Liao, H. Y. M. (2001), 'Multipurpose watermarking for image authentication and protection', *Image Processing, IEEE Transactions on* **10**(10), 1579–1592.

Lu, C. S. and Liao, H. Y. M. (2003), Structural digital signature for image authentication: an incidental distortion resistant scheme, *in* 'IEEE Trans. actions on Multimedia', Vol. 5, pp. 161–173.

Lu, H., Shen, R. and Chung, F. L. (2003), 'Fragile watermarking scheme for image authentication', *Electronics Letters* **39**(12), 898–900.

Lv, X. and Wang, Z. J. (2008), Fast Johnson-Lindenstrauss Transform for robust and secure image hashing, *in* 'Multimedia Signal Processing, 2008 IEEE 10th Workshop on', pp. 725–729.

Lv, X. and Wang, Z. J. (2009), 'An extended image hashing concept: content-based fingerprinting using fjlt', *EURASIP J. Inf. Secur.* **2009**, 2:1–2:16.
URL: *http://dx.doi.org/10.1155/2009/859859*

Mallat, S. G. (1989), 'A theory for multiresolution signal decomposition: the wavelet representation', *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **11**(7), 674–693.

Meixner, A. and Uhl, A. (2005), Security Enhancement of Visual Hashes Through Key Dependent Wavelet Transformations, *in* 'Image Analysis and Processing ICIAP 2005',

Vol. 3617 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 543–550.

Meixner, A. and Uhl, A. (2006), Robustness and security of a wavelet-based CBIR hashing algorithm., *in* 'In Proceeding of the 8th ACM Workshop om Multimedia and Security', pp. 140–145.

Mihçak, K. and Venkatesan, R. (2001), A perceptual audio hashing algorithm: A Tool for Robust Audio Identification and Information Hiding, *in* 'Information Hiding', pp. 51–65.

Mihçak, M. K. and Venkatesan, R. (2002), New iterative Geometric Methods for Robust Perceptual Image Hashing, *in* 'Revised Papers from the ACM CCS-8 Workshop on Security and Privacy in Digital Rights Management', DRM '01, Springer-Verlag, London, UK, pp. 13–21.

Monga, V. (2005), Perceptually Based Methods for Robust Image Hashing, Ph.D thesis, University of Texas.

Monga, V. and Evans, B. L. (2004), Robust perceptual image hashing using feature points, *in* 'Image Processing, 2004. ICIP '04. 2004 International Conference on', Vol. 1, pp. 677–680 Vol. 1.

Monga, V. and Evans, B. L. (2006), Perceptual image hashing via feature points:performance evaluation and tradeoffs, *in* 'IEEE Trans. on Image Process.', Vol. 15, pp. 3453–3466.

Monga, V. and Mihçak, M. K. (2007), Robust and Secure Image Hashing via Non-Negative Matrix Factorizations, *in* 'IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY', Vol. 2, pp. 376–390.

Monga, V., Vats, D. and Evans, B. L. (2005), Image Authentication Under Geometric Attacks Via Structure Matching, *in* 'Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on', pp. 229–232.

Motwani, R. and Raghavan, P. (1996), *Randomized Algorithms*, Cambridge University Press.

Mucedero, A., Lancini, R. and Mapelli, F. (2004), A novel hashing algorithm for video sequences, *in* 'Image Processing, 2004. ICIP '04. 2004 International Conference on', Vol. 4, pp. 2239–2242 Vol. 4.

Norcen, R. and Uhl, A. (2005), 'Robust visual hashing using jpeg 2000', **175**, 223–235.

Patrice, A. (1997), *Ondelettes et turbulances,DIDEROT EDITEUR, ARTS ET SCIENCES*, Paris.

Pennebaker, W. B. and Mitchell, J. L. (1992), *JPEG Still Image Data Compression Standard*, 1st edn, Kluwer Academic Publishers, Norwell, MA, USA.

Pentti, P. (1997), 'Least squares formulation of robust non-negative factor analysis', *Chemometrics and Intelligent Laboratory Systems* **37**(1), 23 – 35.
URL: *http://www.sciencedirect.com/science/article/pii/S0169743996000445*

Ponomarenko, N. N., Bazhyna, A. V. and Egiazarian, K. O. (2007), Prediction of signs of dct coefficients in block-based lossy image compression., *in* J. Astola, K. O. Egiazarian and E. R. Dougherty, eds, 'IPAS', Vol. 6497 of *SPIE Proceedings*, SPIE, p. 64970.

Prungsinchai, S., Khelifi, F. and Bouridane, A. (2012), Sub-images based on image hashing with Non-Negative Matrix Factorization, *in* 'The 19th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2012', pp. 781–784.

Prungsinchai, S., Khelifi, F. and Bouridane, A. (2013), Foureie-Mellin Transform for Robust Image Hashing, *in* 'The 4th International Conference on Emerging Security Technologies (EST), 2013', pp. 58–61.

Roover, D. C., Vleeschouwer, D. C., Lefebvre, F. and Macq, B. (2005), 'Robust video hashing based on radial projections of key frames', *Signal Processing, IEEE Transactions on* **53**(10), 4020–4037.

Roy, S. and Sun, Q. (2007), Robust hash for detecting and localizing image tampering, *in* 'Image Processing, 2007. ICIP 2007. IEEE International Conference on', Vol. 6, pp. VI – 117–VI – 120.

Roy, S., Sun, Q. and Kalker, T. (2008), Performance analysis of locality preserving image hash, *in* 'Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on', pp. 1268–1271.

Roy, S., Zhu, X., Yuan, J. and Chang, E. C. (2007), 'On preserving robustness-false alarm tradeoff in media hashing'.
URL: *http://dx.doi.org/10.1117/12.704837*

Schneider, M. and Chang, S. F. (1996), A robust content based digital signature for image authentication, *in* 'Proc. IEEE conf. on image processing', Vol. 3, pp. 227–230.

Seo, S. J., Haitsma, J., Kalker, T. and Yoo, D. C. (2004), 'A robust image fingerprinting system using the Radon transform', *Signal Processing: Image Communication* **19**(4), 325 – 339.
URL: *http://www.sciencedirect.com/science/article/pii/S0923596503001541*

Smith, M. J. T. and Barnwell, T. P. (1987), 'A new filter bank theory for time-frequency representation', *Acoustics, Speech and Signal Processing, IEEE Transactions on* **35**(3), 314–327.

Sun, Q., Tian, Q. and Chang, S. F. (2002), A robust and secure media signature scheme for jpeg images, *in* 'Multimedia Signal Processing, 2002 IEEE Workshop on', pp. 296–299.

Sunil, L. and Yoo, C. D. (2008), 'Robust Video Fingerprinting for Content-Based Video Identification', *Circuits and Systems for Video Technology, IEEE Transactions on* **18**(7), 983–988.

Swaminathan, A., Mao, Y. and Wu, M. (2004), Image hashing resilient to geometric and filtering operations, *in* 'Multimedia Signal Processing, 2004 IEEE 6th Workshop on', pp. 355–358.

Swaminathan, A., Mao, Y. and Wu, M. (2006), 'Robust and secure image hashing', *Information Forensics and Security, IEEE Transactions on* **1**(2), 215–230.

Swets, J. A., Dawer, R. M. and Monahan, J. (2000), Better decisions through science, *in* 'Scientific American', Vol. 283, USA.

Tang, S., Li, J. T. and Zhang, Y. D. (2005), Compact and Robust Image Hashing., *in* 'ICCSA (2)', Vol. 3481 of *Lecture Notes in Computer Science*, Springer, pp. 547–556.

Tang, Z., Dai, Y., Zhang, X. and Zhang, S. (2012), Perceptual image hashing with histogram of color vector angles, *in* 'Proceedings of the 8th international conference on Active Media Technology', Springer-Verlag, Berlin, Heidelberg, pp. 237–246.

Tang, Z., Wang, S., Zhang, X., Wei, W. and Su, S. (2008), Robust image hashing for tamper detection using non-negative matrix factorization, Vol. 2, pp. 18–26.

Tang, Z., Zhang, X. and Zhang, S. (2013), Robust Perceptual Image Hashing Based on Ring Partition and NMF, *in* 'IEEE Transactions on Knowledge and Data Engineering', Vol. 99, Los Alamitos, CA, USA.

Vaidyanathan, P. P. (1987), 'Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques', *ASSP Magazine, IEEE* **4**(3), 4–20.

Venkatesan, R., Koon, S. M., Jakubowski, M. H. and Moulin, P. (2000), Robust image hashing, *in* 'Proc. IEEE conf. on image processing', Vol. 3, pp. 664–666.

Vetterli, M. (1987), 'A theory of multirate filter banks', *Acoustics, Speech and Signal Processing, IEEE Transactions on* **35**(3), 356–372.

Villasenor, J. D., Belzer, B. and Liao, J. (1995), 'Wavelet filter evaluation for image compression', *Image Processing, IEEE Transactions on* **4**(8), 1053–1060.

Wu, M. and Liu, B. (1998), Watermarking for image authentication, *in* 'Proc. IEEE Conf. on Image Processing', Vol. 2, pp. 437–477.

Wu, M. N., Lin, C. C. and Chang, C. C. (2007), 'Novel image copy detection with rotating tolerance', *Journal of Systems and Software* **80**(7), 1057 – 1069.

Xiang, S., Kim, H. and Huang, J. (2007), 'Histogram-based image hashing scheme robust against geometric deformations', *The 9th ACM Multimedia and Security Workshop* pp. 121–128.

Xiang, S., Yang, J. and Huang, J. (2012), 'Perceptual video hashing robust against geometric distortions', *Science China Information Sciences* **55**(7), 1520–1527.

Xie, L. and Arce, G. R. (2001), A class of authentication digital watermarks for secure multimedia communication, *in* 'IEEE Trans.on Image Processing', Vol. 10, S1057-7149(01)09362-9, pp. 1754–1756.

Xie, L., Arce, G. R. and Graveman, R. F. (2001), 'Approximate image message authentication codes', *Trans. Multi.* **3**(2), 242–252.

Yang, O. and Rhee, K. H. (2010), A survey on image hashing for image authentication, number 5, pp. 1020–1030.

Yeung, M. M. and Mintzer, F. (1997), An invisible watermarking scheme for image verification, *in* 'Proc. IEEE Conf. on Image Processing', Vol. 1, pp. 680–683.

Yu, L. and Sun, S. (2006), Image Robust Hashing Based on DCT Sign, *in* 'Intelligent Information Hiding and Multimedia Signal Processing, 2006. IIH-MSP '06. International Conference on', pp. 131–134.

Zauner, C. (2012), Implementation and Benchmarking of Perceptual Image Hash Functions.

Zou, F., Ling, H., Li, X., Xu, Z. and Li, P. (2009), Robust image copy detection using local invariant feature, *in* 'Multimedia Information Networking and Security, 2009. MINES '09. International Conference on', Vol. 1, pp. 57–61.

Zweig, X. H. and Campbell, G. (1993), Reciever-operating characteristic (roc) plot: a fundamental evaluation tool in clinical medicine, *in* 'Clinical Chemistry', pp. 561–577.